

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 11, 2004, 15:58:00 ; Search time 61 Seconds

(without alignments)  
7605.865 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923  
Sequence: 1 MASPATGVLPFPPLLL.....AGTVDSGSGEPFLPLH 2923Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:\*  
23: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:\*  
24: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match length	DB ID	Description
1	2923	100.0	2923	22	AAU07054 Human Flamingo pro
2	2923	100.0	2923	23	AAU50866 Cadherin EGF LAG s
3	2923	100.0	2923	24	ABP82018 Human GPCR CELSR2
4	2894	99.0	2894	24	ABU11556 Human MDT1 polypep
5	2837	97.1	2956	22	AAU07053 Human Flamingo pol
6	1958	67.0	2936	23	AAU74826 Human REPT9 prot
7	1917	65.6	2560	22	ABU11404 Human FLAMINGO 1 h
8	1623	55.5	2405	21	AAU842192 Human ORFX ORF1956
9	534	18.3	568	19	AAU51244 Human Calcitonin r

10	485	16.6	565	24	ABJ37074 Human breast cance
11	281	9.6	568	21	ABP89665 Human polypeptide
12	281	9.6	717	21	AAU56721 Human prostate can
13	220	7.5	470	24	ABP75877 Human secretory po
14	118	4.0	219	22	ABP03070 Human expressed po
15	118	4.0	219	22	AAU19681 Human novel extrac
16	118	4.0	219	22	ABP47901 Human polypeptide
17	100	3.4	111	22	ABU17070 Human nervous syst
18	100	3.4	111	22	AAU19829 Human novel extrac
19	100	3.4	111	22	ABP48049 Human polypeptide
20	82	2.8	96	22	ABU10351 Human CDNA SEQ ID
21	82	2.8	96	22	AAU18432 Human novel uterin
22	82	2.8	96	22	AAU17045 Human novel secret
23	82	2.8	96	22	AAU19897 Human novel calcin
24	82	2.8	96	22	ABP66938 Human polypeptide
25	82	2.8	96	22	ABU05759 Human human protei
26	42	1.4	141	22	AAU89890 Human immune/haema
27	42	1.4	141	22	AAU19685 Human novel extrac
28	42	1.4	141	22	ABP47905 Human polypeptide
29	23	0.8	2707	18	AAU27161 Mouse receptor ME2
30	21	0.7	44	17	AAU87108 Rat-316 cadherin p
31	21	0.7	45	15	AAU85866 Human GPCR CELSR1
32	20	0.7	20	23	AAU50874 G protein-coupled
33	20	0.7	20	24	ABP83213 G protein-coupled
34	20	0.7	20	24	ABP83214 G protein-coupled
35	20	0.7	3014	22	AAU68533 Human novel cytol
36	20	0.7	3014	22	AAU02196 Seven-pass transme
37	20	0.7	3014	24	ABP81979 Human GPCR CELSR1
38	20	0.7	3028	22	AAU08586 Human NOV7 protein
39	19	0.7	19	24	ABP83212 G protein-coupled
40	19	0.7	19	24	ABP83215 G protein-coupled
41	18	0.6	18	23	AAU50873 Cadherin EGF LAG s
42	18	0.6	18	24	ABP83216 G protein-coupled
43	18	0.6	874	18	AAU07600 Human G-protein re
44	18	0.6	874	20	AAU94075 Human G-protein co
45	18	0.6	884	19	AAU59666 Amino acid sequenc

## ALIGNMENTS

RESULT 1	AAU07054	AAU07054 standard; Protein; 2923 AA.
AC	AAU07054;	
XX		
XX	24-OCT-2001 (first entry)	
DT		
XX		Human Flamingo protein encoded by cDNA splice variant.
XX		
KW	Flamingo; human; splice variant; G-protein coupled receptor; diabetes;	
KW	signal transduction pathway; bacterial; fungal; protozoan; cancer;	
KW	anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;	
KW	obesity; hypotension; hypertension; urinary retention; angina pectoris;	
KW	myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;	
KW	benign prostatic hypertrophy; psychotic disorder; neurological disorder;	
KW	manic depression; delirium; dementia; severe mental retardation;	
KW	Huntington's disease; Gilles de la Tourette's syndrome; anorectic;	
KW	antifungal; antiviral; antiprotozoal; anti-HIV; anorectic; anti-anorectic;	
KW	antiparkinsonian; cardiac; cerebroprotective; neuroprotective;	
KW	antidepressant; anticonvulsant; antitense therapy; gene therapy.	
OS	Homo sapiens.	
XX		
PN	MO200161003-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	19-FEB-2001; 2001WO-GB00680.	
XX		
PR	19-FEB-2000; 2000GB-0004196.	
XX		

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX Testa TT;  
 XX WPI: 2001-502792/55.  
 DR N-PSDB; AAS11678.  
 PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 XX  
 PS Claim 2; Page 33-34; 66pp; English.  
 CC The sequence represents a human Flamingo polypeptide encoded by a cDNA  
 CC splice variant polynucleotide. Flamingo is a member of the G-protein  
 CC coupled receptor family, which is involved in signal transduction  
 CC pathways. By screening to identify compounds that stimulate or inhibit  
 CC the function or level of the protein, treatments can be developed for  
 CC various diseases and bacterial, fungal, protozoan and viral infections,  
 CC including HIV, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, angina pectoris, myocardial infarction, stroke,  
 CC ulcers, allergies and benign prostatic hypertrophy. Also treatable are  
 CC psychotic and neurological disorders such as anxiety, schizophrenia,  
 CC manic depression, delirium, dementia, severe mental retardation,  
 CC Huntington's disease and Gilles de la Tourette's syndrome.  
 CC  
 SQ Sequence 2923 AA;  
 Query Match 100.0%; Score 2923; DB 22; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVVTVSAVDRDAHSVITVQITSGNTRNRSFITS 660  
 DB 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVVTVSAVDRDAHSVITVQITSGNTRNRSFITS 660  
 QY 661 QSGGGVSLALPLDYTLERQVLAATASTGTDQTNQIVVNTDANTHRPVQSSHYTN 720  
 DB 661 QSGGGVSLALPLDYTLERQVLAATASTGTDQTNQIVVNTDANTHRPVQSSHYTN 720  
 QY 721 VNEDRPAGTTVVLISATDDEGTENARITFEMEDSIPOFRIADPTGAVTTQAEIDYEDQVS 780  
 DB 721 VNEDRPAGTTVVLISATDDEGTENARITFEMEDSIPOFRIADPTGAVTTQAEIDYEDQVS 780  
 QY 781 YTLAATARDNGIPQKSDTTYLSILVNDVNDNAPQFLRDSYQGSVEYDVPFTSVLQISAT 840  
 DB 781 YTLAATARDNGIPQKSDTTYLSILVNDVNDNAPQFLRDSYQGSVEYDVPFTSVLQISAT 840  
 QY 841 DRDSGLNGVFTVTPQGGDGDGDFVSTSGVTRTLRLDRENVAYVLAAYADGMP 900  
 DB 841 DRDSGLNGVFTVTPQGGDGDGDFVSTSGVTRTLRLDRENVAYVLAAYADGMP 900  
 QY 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVATDPDEGTNAQIMYQI 960  
 DB 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVATDPDEGTNAQIMYQI 960  
 QY 961 VEGNIPEVFQLDIFSGELTALVDLYEDRPEYVLVIQATSAPLVSATVVRLLDRNDP 1020  
 DB 961 VEGNIPEVFQLDIFSGELTALVDLYEDRPEYVLVIQATSAPLVSATVVRLLDRNDP 1020  
 QY 1021 PVLGNFELLFNNYVNNRSPFGGAI GRVAPADPOISLSTYSPFENGSLVTLNASTG 1080  
 DB 1021 PVLGNFELLFNNYVNNRSPFGGAI GRVAPADPOISLSTYSPFENGSLVTLNASTG 1080  
 QY 1081 ELKLSRALDNNRPLEAIMSVLSDGVHSTAAQALRVITITTEMLTHSTITLEDMSPER 1140  
 DB 1081 ELKLSRALDNNRPLEAIMSVLSDGVHSTAAQALRVITITTEMLTHSTITLEDMSPER 1140  
 QY 1141 FLSPILGLFTQVAATLATPDPHVVFVYQDRTDAPGHIILNVSLVSGPPGPGGPPYL 1200  
 DB 1141 FLSPILGLFTQVAATLATPDPHVVFVYQDRTDAPGHIILNVSLVSGPPGPGGPPYL 1200  
 QY 1201 PSEDOERLYLNRSLTALSAQVLPEDNITCLREPCENYMCVSLRDPSSAPFASS 1260  
 DB 1201 PSEDOERLYLNRSLTALSAQVLPEDNITCLREPCENYMCVSLRDPSSAPFASS 1260  
 QY 1261 VLFRRPHVPGRLRCRCPPEFTGDCYCEVDLYCYSRPCGPHGRCSREGGYTCLCRDGYTG 1320  
 DB 1261 VLFRRPHVPGRLRCRCPPEFTGDCYCEVDLYCYSRPCGPHGRCSREGGYTCLCRDGYTG 1320  
 QY 1321 EHCEVSARSGRCTPGVCNKGTCVNLVGGFKDCPSGDPPEKPYCQVTRSPFASFTIF 1380  
 DB 1321 EHCEVSARSGRCTPGVCNKGTCVNLVGGFKDCPSGDPPEKPYCQVTRSPFASFTIF 1380  
 QY 1381 RGLRORFHFTLASFTEKRDGLLYNGRPNKHDPVLALEVOVOLTFPSAGESTTYS 1440  
 DB 1381 RGLRORFHFTLASFTEKRDGLLYNGRPNKHDPVLALEVOVOLTFPSAGESTTYS 1440  
 QY 1441 PFVPGVSDGOMHTVQLKTYNPLLGQTGLPOGSPSEKQAVVTVVDCGDTGVALRFGSVLG 1500  
 DB 1441 PFVPGVSDGOMHTVQLKTYNPLLGQTGLPOGSPSEKQAVVTVVDCGDTGVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQGSKSLDTGPLLGGVDPDLBESFVRROROFQCMKRNQOVSRHIDMAD 1560  
 DB 1501 NYSCAAQGTQGSKSLDTGPLLGGVDPDLBESFVRROROFQCMKRNQOVSRHIDMAD 1560  
 QY 1561 FIANNGTGCPAKKXVCNSNTCHNGTGVNOMDAPSCCPJGFGKSCAOEMANPOHL 1620  
 DB 1561 FIANNGTGCPAKKXVCNSNTCHNGTGVNOMDAPSCCPJGFGKSCAOEMANPOHL 1620  
 QY 1621 GSSLVAMHGLSLPISQWTLSTMFRTROADGVLLQITRGRSTITTLQREGHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISQWTLSTMFRTROADGVLLQITRGRSTITTLQREGHVMSVEGT 1680  
 QY 1681 GLQASSLRLEPGRANDGMHHAQALAGAGGFGHALISFDYGOQARAGNUGPRLHGLHS 1740



```

Db      1681  GLASSLRLEPGANODMWHADLALGASGPGHALLSPYGOORABGNLGPRLAGLALS 1740
Qy      1741  NITVGGI PGAGVAGFRGCLQGVRSVTPEGVNSIDPSHGESINVEQCSLPDPDSN 1800
Db      1741  NITVGGI PGAGVAGFRGCLQGVRSVTPEGVNSIDPSHGESINVEQCSLPDPDSN 1800
Qy      1801  PCFANYSAMDMSYSCSDPGYGDNCINVCOLNPEHOSVCTKRPSAHGTYTCBPEN 1860
Db      1801  PCFANYSAMDMSYSCSDPGYGDNCINVCOLNPEHOSVCTKRPSAHGTYTCBPEN 1860
Qy      1861  YLGPYETRIDQCPRGMMHPTCGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSP 1920
Db      1861  YLGPYETRIDQCPRGMMHPTCGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSP 1920
Qy      1921  CLLDCYPTGSLSRVCDPEDGQCPCKPGVIGROCDRCDNPFAVTTNGCEVNYDSCE 1980
Db      1921  CLLDCYPTGSLSRVCDPEDGQCPCKPGVIGROCDRCDNPFAVTTNGCEVNYDSCE 1980
Qy      1981  EAGIMWPRTRFGI PAAPCPKGSFGTAVRHCDEHRCGLPNNLNCITSITSELSKFAERL 2040
Db      1981  EAGIMWPRTRFGI PAAPCPKGSFGTAVRHCDEHRCGLPNNLNCITSITSELSKFAERL 2040
Qy      2041  QRNESGLDSGRSQQALLLNATQHTAGYFGSDVKVAYQATRLAHSTORGEGLSATQ 2100
Db      2041  QRNESGLDSGRSQQALLLNATQHTAGYFGSDVKVAYQATRLAHSTORGEGLSATQ 2100
Qy      2101  DVHFTENLRVGSALLDTANKRMWELIQTEGGTAMLOHYEAYASALANMBHTYLSPP 2160
Db      2101  DVHFTENLRVGSALLDTANKRMWELIQTEGGTAMLOHYEAYASALANMBHTYLSPP 2160
Qy      2161  TITPNIIVISVRLDKNFAGACLPRYEALRGROPPLFTTVILPESVFETPPVVRPAG 2220
Db      2161  TITPNIIVISVRLDKNFAGACLPRYEALRGROPPLFTTVILPESVFETPPVVRPAG 2220
Qy      2221  PGSAORPEELARQRHPELSQGEAAVSVIITRTLAGLPHNTPDPGRSLRVKRPINT 2280
Db      2221  PGSAORPEELARQRHPELSQGEAAVSVIITRTLAGLPHNTPDPGRSLRVKRPINT 2280
Qy      2281  PVSIVSVHDEBELPRALDKPVTVOFPLETETERTKPCVFMNHSILVSGTGMARSCE 2340
Db      2281  PVSIVSVHDEBELPRALDKPVTVOFPLETETERTKPCVFMNHSILVSGTGMARSCE 2340
Qy      2341  VVRNRSNHSVCCNMHTSPAVLMDVSRRENGELLPLKTLTYVALGTLALTLTFFPLT 2400
Db      2341  VVRNRSNHSVCCNMHTSPAVLMDVSRRENGELLPLKTLTYVALGTLALTLTFFPLT 2400
Qy      2401  LRILRSNQHGRIRNLTALGLAQVFLAGINQADLPACTVAILLHFLYLCFTFSNALLE 2460
Db      2401  LRILRSNQHGRIRNLTALGLAQVFLAGINQADLPACTVAILLHFLYLCFTFSNALLE 2460
Qy      2461  ALHLRYALTEVRDVTGPMFYYTLMGVPATFTGLAVGLDPBGYGNPFCWLSIYDTLI 2520
Db      2461  ALHLRYALTEVRDVTGPMFYYTLMGVPATFTGLAVGLDPBGYGNPFCWLSIYDTLI 2520
Qy      2521  WSRAGVAVRVSNSVFLYTLAARASCAHQOQFEKGPVSGLOPSRAVLLLSATWLLAL 2580
Db      2521  WSRAGVAVRVSNSVFLYTLAARASCAHQOQFEKGPVSGLOPSRAVLLLSATWLLAL 2580
Qy      2581  LSVNSSTLLFHYLPATCNCIQGFPIFASVYVLSKEYRKALKACSKRPSPDPLATKSTL 2640
Db      2581  LSVNSSTLLFHYLPATCNCIQGFPIFASVYVLSKEYRKALKACSKRPSPDPLATKSTL 2640
Qy      2641  TSSYNCPSPYADRLYQPYDASGSLHSTRSKSGOPSYIPFLIRBSALNPGQPGGLG 2700
Db      2641  TSSYNCPSPYADRLYQPYDASGSLHSTRSKSGOPSYIPFLIRBSALNPGQPGGLG 2700
Qy      2701  DPGSLLEGGDOOHDPDTSDDLLEDDOSGYASTHSDESEBBEREKAAPFEGEG 2760
Db      2701  DPGSLLEGGDOOHDPDTSDDLLEDDOSGYASTHSDESEBBEREKAAPFEGEG 2760
Qy      2761  WBSILPGAEARLPLHSTPKDGGPGKAPWPGFTTAKSSNGNAPBEERLRNGDALSR 2820

```

```

Db      2761  WBSILPGAEARLPLHSTPKDGGPGKAPWPGFTTAKSSNGNAPBEERLRNGDALSR 2820
Qy      2821  EBSLGLPSSAQPXKGLIKKCLPTISEKSLRLPLEQCTGSSRSSASBEGSGCPP 2880
Db      2821  EBSLGLPSSAQPXKGLIKKCLPTISEKSLRLPLEQCTGSSRSSASBEGSGCPP 2880
Qy      2881  RPPRQSLQEQNGWPIAMSTIKAGTVNDDSSGSEPLFPNPLH 2923
Db      2881  RPPRQSLQEQNGWPIAMSTIKAGTVNDDSSGSEPLFPNPLH 2923

RESULT 2
AAMS0866
ID AAMS0866 standard; Protein; 2923 AA.
XX
AC AAMS0866;
XX
DT 07-MAY-2002 (first entry)
XX
DE Cadherin EGF LAG seven-pass G-type receptor 2, basal cell marker.
XX
KW Cadherin EGF LAG seven-pass G-type receptor 2; receptor;
KW EGF-like domain, multiple 2; basal cell; marker; breast cancer;
KW tumour; diagnosis; prognosis; classification; human; gene therapy.
XX
OS Homo sapiens.
XX
PN W0200208765-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US23843.
XX
PR 26-JUL-2000; 2000US-220967P.
XX
PA (STRD) UNIV STANFORD.
PA (GENO-) APPLIED GENOMICS INC.
XX
PI Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
PI Seltz R;
XX
DR WPI; 2002-206094/26.
XX
PT Classifying and treating breast tumours associated with the expression
PT and/or cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like
PT Domain, Multiple 2 -
XX
PS Claim 3; Fig 1C; 741pp; English.
XX
CC The present sequence is that of human cadherin EGF LAG seven-pass
CC G-type receptor 2, also known as EGF-like domain, multiple 2.
CC The invention provides new reagents and methods for the management
CC (e.g. detection, classification, provision of diagnostic and
CC prognostic information, treatment, etc.) of breast cancer. CDNA
CC microarray technology was used to identify genes whose expression
CC profile across a large group of tumour samples correlated with that
CC of cytokeratin 5 and cytokeratin 17, markers for basal cells of the
CC normal mammary lactation gland. Tumours that express cytokeratin
CC 5/6 and/or 17 have a poor prognosis relative to tumours overall.
CC The basal marker genes identified included those encoding cadherin
CC 3 (see AAMS0864), matrix metalloproteinase 14 (see AAMS0865) and
CC cadherin EGF LAG seven-pass G-type receptor 2. Detection of these
CC markers allows classification of a tumour sample as belonging to a
CC basal tumour subclass. By providing reagents that may reliably be
CC used to classify tumours as belonging to a basal subclass, the
CC invention enables a variety of methods for improving therapeutic
CC options for patients with breast cancer and for individualising
CC therapy. Potential therapeutic agents include compounds that
CC modulate breast basal cell marker genes or that modulate
CC polypeptides encoded by these genes. In particular, therapeutic
CC agents include antibodies that specifically bind to the basal
CC marker proteins and which recognise basal cells of normal mammary

```



1981 EAGIWMPTRTFGLPAAPCPKSGRTAVRHCDHHRGMLPMLFVCTSTTSEKGPAREL 2040  
 QY 2041 QNBSGIDGSRSDQALLLNATQHTAGYFGSDVKAAYQATATLLAES7QORGESATQ 2100  
 Db 2041 QNBSGIDGSRSDQALLLNATQHTAGYFGSDVKAAYQATATLLAES7QORGESATQ 2100  
 QY 2101 DVHPTTELAVGSAALLDTANKRHHELLQOEBGCTAMLLQHYEAVASLAQNMHTTYSF 2160  
 Db 2101 DVHPTTELAVGSAALLDTANKRHHELLQOEBGCTAMLLQHYEAVASLAQNMHTTYSF 2160  
 QY 2161 TIVTPNIVISVRLDKGNFAGAKLPYEALRGBOPLLETTVLLPESVPRETPPVAPAG 2220  
 Db 2161 TIVTPNIVISVRLDKGNFAGAKLPYEALRGBOPLLETTVLLPESVPRETPPVAPAG 2220  
 QY 2221 PGEAQBEEELARRORRHELPESGEAVASVIIYRTLAGLPHNYDPDKRSRVPKRPINT 2280  
 Db 2221 PGEAQBEEELARRORRHELPESGEAVASVIIYRTLAGLPHNYDPDKRSRVPKRPINT 2280  
 QY 2281 PVVSIHVHDEBELPRALDKPVTVQFRLTEERTKPICTPMNHSILVSGTGMASRGCE 2340  
 Db 2281 PVVSIHVHDEBELPRALDKPVTVQFRLTEERTKPICTPMNHSILVSGTGMASRGCE 2340  
 QY 2341 VVERNSHVSQCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFFPLTL 2400  
 Db 2341 VVERNSHVSQCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFFPLTL 2400  
 QY 2401 LRLRSNQHGRIRNLTALGLAOLVFLLAGINADLPACTVIALHFLVLCFESWALTE 2460  
 Db 2401 LRLRSNQHGRIRNLTALGLAOLVFLLAGINADLPACTVIALHFLVLCFESWALTE 2460  
 QY 2461 ALHLVALTEVPDVTNGRPFYTMGLGWVPAFTTGLAVGLDPGKGNPDCMTSTYDTLI 2520  
 Db 2461 ALHLVALTEVPDVTNGRPFYTMGLGWVPAFTTGLAVGLDPGKGNPDCMTSTYDTLI 2520  
 QY 2521 WSPAGPVAFAVMSVFLYIILAAASCAAOQGEFKGPGVGLPSPAVLILLSATWLLAL 2580  
 Db 2521 WSPAGPVAFAVMSVFLYIILAAASCAAOQGEFKGPGVGLPSPAVLILLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHYLPATNCIOGPFILSYVLSKEVAKALKACSRKSPDPALTTKSTL 2640  
 Db 2581 LSVNSDTLLFHYLPATNCIOGPFILSYVLSKEVAKALKACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNSPSPADRLVOPYGDASGSLHSTRSGKSQSYIPLILRESALNPGQGPGLG 2700  
 Db 2641 TSSYNSPSPADRLVOPYGDASGSLHSTRSGKSQSYIPLILRESALNPGQGPGLG 2700  
 QY 2701 DPGSLFLEGQDQHDPTDSDLSLEDDQSGSYASTHSSDESEEESEEEAAPFGEQ 2760  
 Db 2701 DPGSLFLEGQDQHDPTDSDLSLEDDQSGSYASTHSSDESEEESEEEAAPFGEQ 2760  
 QY 2761 WDSLCPGABRLPLHSTPKDGFPGKAPWPGDFTTAKSSGNGAPBEERLRENGDALSR 2820  
 Db 2761 WDSLCPGABRLPLHSTPKDGFPGKAPWPGDFTTAKSSGNGAPBEERLRENGDALSR 2820  
 QY 2821 EGSILGPGSSAPPHGILLKKKLLPTISEKSLILPLRECTSSSGSSASBSRGGPP 2880  
 Db 2821 EGSILGPGSSAPPHGILLKKKLLPTISEKSLILPLRECTSSSGSSASBSRGGPP 2880  
 QY 2881 RPPPROSLOEQLNGVPIAMSIKAGTVDEDSGSEPLFFNFLLH 2923  
 Db 2881 RPPPROSLOEQLNGVPIAMSIKAGTVDEDSGSEPLFFNFLLH 2923

KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer.  
 OS Homo sapiens.  
 PN WO200261087-A2.  
 XX 08-AUG-2002.  
 PD 19-DEC-2001; 2001MO-US50107.  
 PF 19-DEC-2000; 2000US-257144P.  
 PR 19-DEC-2000; 2000US-257144P.  
 PA (LIFE-) LIFESPAN BIOSCIENCES INC.  
 XX Burner GC, Roush CL, Brown JP,  
 XX WPI; 2003-046718/04.  
 DR N-PSDB; AB242868.  
 XX  
 PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating  
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
 PT cancer or autoimmune diseases  
 PS Disclosure; Fig 1; 523pp; English.  
 XX  
 CC The present invention describes antigenic peptides (I) comprising:  
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular  
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity  
 CC or avidity for a particular GPCR. (1) can be used as GPCR modulators and  
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
 CC an antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related diseases, growth-related diseases, cell  
 CC regeneration-related disease, immunological-related cell proliferative  
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
 CC hypotension, renal disorder, rheumatoid arthritis, trauma, ulcers, or  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 2923 AA;  
 Query Match 100.0%; Score 2923; DB 24; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;  
 Matches 2923; Conservative 0; Indels 0; Gaps 0;  
 QY 1 MSAPATGVPLPTPPPLLLLLLLPPPLLDQVGPCCSLGSRGSSGACAPMGWLCPS 60  
 Db 1 MSAPATGVPLPTPPPLLLLLLLPPPLLDQVGPCCSLGSRGSSGACAPMGWLCPS 60  
 QY 61 SASNMLTTSRCRDAGTETLGHVPHHDLRWCEBSEAHITLPPAPBGCPSMCLLIG 120

Dh	61	SASINMLTYSRCRAGTELTHGLVPHHDLRWCPBESNAHPLPAPAPGCGWSCRLLGIG	120
Qy	121	GHLSPOGKLTLPBEHPCLAKPRLRCQCKLAQAPGLRAGERSPEPSLGRKREKNAVNTAPQ	180
Dh	121	GHLSPOGKLTLPBEHPCLAKPRLRCQCKLAQAPGLRAGERSPEPSLGRKREKNAVNTAPQ	180
Qy	181	POPSYQATVPENOPAGTPVASLRAIDDEBAGALEYTMALPDSRSNOFPSLDPTVGA	240
Dh	181	POPSYQATVPENOPAGTPVASLRAIDDEBAGALEYTMALPDSRSNOFPSLDPTVGA	240
Qy	241	VTTBEIDRETKSTHVFVTAQDHGMPRSALATLTITVDTNHDHPFEQOEKESLRE	300
Dh	241	VTTBEIDRETKSTHVFVTAQDHGMPRSALATLTITVDTNHDHPFEQOEKESLRE	300
Qy	301	NLEVEGYEVLTVRATDGDAPPNANILYRLLESGSGSPSEVEFIDPRSGLVTRGVNDEEV	360
Dh	301	NLEVEGYEVLTVRATDGDAPPNANILYRLLESGSGSPSEVEFIDPRSGLVTRGVNDEEV	360
Qy	361	ESYQUTVEASDQGDHPPGRSTTAAVFLSVEDNDNAPQFSKRYVQVREDDVTGAPLKR	420
Dh	361	ESYQUTVEASDQGDHPPGRSTTAAVFLSVEDNDNAPQFSKRYVQVREDDVTGAPLKR	420
Qy	421	VTASDRKGSNAVVHYSIMSGNARGQFYLDQGTALDVSPLDYETKEVTLVRADQGG	480
Dh	421	VTASDRKGSNAVVHYSIMSGNARGQFYLDQGTALDVSPLDYETKEVTLVRADQGG	480
Qy	481	RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVHQAIDADAGDNARL	540
Dh	481	RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVHQAIDADAGDNARL	540
Qy	541	EYRLAGVGHDPFPTINNGTGMISVAABLDREEDVDFSGVEARHDGHPALASASVTV	600
Dh	541	EYRLAGVGHDPFPTINNGTGMISVAABLDREEDVDFSGVEARHDGHPALASASVTV	600
Qy	601	LDVNDNPTFOPEYTVRLNEDAAVGTSVTVSAVNDASHVITVOYTGSTRRPSITS	660
Dh	601	LDVNDNPTFOPEYTVRLNEDAAVGTSVTVSAVNDASHVITVOYTGSTRRPSITS	660
Qy	661	QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVNTDANTHRPVFGSHHTVN	720
Dh	661	QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVNTDANTHRPVFGSHHTVN	720
Qy	721	VNERPAGTIVVILSADDEOTGENARITTFMEDSIPOFRIDADTGAVTTOAELEDEQVS	780
Dh	721	VNERPAGTIVVILSADDEOTGENARITTFMEDSIPOFRIDADTGAVTTOAELEDEQVS	780
Qy	781	YTLAITARDNGIPKQSDTTYLEILVNDVNDNAPOFLRDSYQGSYEDVPPTSVLOISAT	840
Dh	781	YTLAITARDNGIPKQSDTTYLEILVNDVNDNAPOFLRDSYQGSYEDVPPTSVLOISAT	840
Qy	841	DRDSGLNGRVEYTFQGGDGDGDPFVBSSTSGIVRTLRLDRENVAAQYVLAAYADKMP	900
Dh	841	DRDSGLNGRVEYTFQGGDGDGDPFVBSSTSGIVRTLRLDRENVAAQYVLAAYADKMP	900
Qy	901	ARTMEYTVTVLVDNDMPVVEODEPVEENSPIGLAVARVATPDBETNAQIMYQI	960
Dh	901	ARTMEYTVTVLVDNDMPVVEODEPVEENSPIGLAVARVATPDBETNAQIMYQI	960
Qy	961	VEGNIPEVFOLDIPSGELTALVDDEDRPREYVIVIOATSAPVSRATVHRLDRNDP	1020
Dh	961	VEGNIPEVFOLDIPSGELTALVDDEDRPREYVIVIOATSAPVSRATVHRLDRNDP	1020
Qy	1021	PVAGNFEILFNNTVYTNRSSSPGGAIGRVPADHDIDSLSLTFSPERGNELSLVLNASTG	1080
Dh	1021	PVAGNFEILFNNTVYTNRSSSPGGAIGRVPADHDIDSLSLTFSPERGNELSLVLNASTG	1080
Qy	1081	ELKISRLLDNNRPLEAITSVYSDGHSYTAQCLARTITIDEMLTHSITLRLDMSPER	1140
Dh	1081	ELKISRLLDNNRPLEAITSVYSDGHSYTAQCLARTITIDEMLTHSITLRLDMSPER	1140
Qy	1141	FLSLPLGLFIOAVAAATLTPPDHVVVFNVDTPDAGGHLNANVSLVGOPPGGCGPPFL	1200
Dh	1141	FLSLPLGLFIOAVAAATLTPPDHVVVFNVDTPDAGGHLNANVSLVGOPPGGCGPPFL	1200
Dh	1201	PSBDLOERLYLANSILTAISAORVLPEDDNI CLRBPCCNMRCSVYRFPSSAFIASSS	1260
Qy	1201	PSBDLOERLYLANSILTAISAORVLPEDDNI CLRBPCCNMRCSVYRFPSSAFIASSS	1260
Dh	1261	VLFRPIHPVGLRCRCBPFGTGDYCETEVDLCYSRPQPHGRCSRREGGYTCLCRDGYTG	1320
Qy	1261	VLFRPIHPVGLRCRCBPFGTGDYCETEVDLCYSRPQPHGRCSRREGGYTCLCRDGYTG	1320
Dh	1321	EHCEVSARSRCRTPGVCKNGTCVNLVGGFKDCDSGDEKXYCQYTTSSFPNHSITP	1380
Qy	1321	EHCEVSARSRCRTPGVCKNGTCVNLVGGFKDCDSGDEKXYCQYTTSSFPNHSITP	1380
Dh	1381	RGLRORPHFTALSPATKERDGLLYNGRPNKHDFTALAEVIOBOVOLTFSAGSTTVS	1440
Qy	1381	RGLRORPHFTALSPATKERDGLLYNGRPNKHDFTALAEVIOBOVOLTFSAGSTTVS	1440
Dh	1441	PFVPGVSDQGMHTVOLKTYNKPDLQYGLPQSPSOKVAVTVTDGCDTVALRFGSVLG	1500
Qy	1441	PFVPGVSDQGMHTVOLKTYNKPDLQYGLPQSPSOKVAVTVTDGCDTVALRFGSVLG	1500
Dh	1501	NYSCLAQGTQSGSKSLDLTGPIILGVPDLPSFPVRMRQFVGCNRLQVDSRHIDMAD	1560
Qy	1501	NYSCLAQGTQSGSKSLDLTGPIILGVPDLPSFPVRMRQFVGCNRLQVDSRHIDMAD	1560
Dh	1561	FIANNGTVPCCPKKVVCSNTECHNGTCVNOQDASCECPFLFGGSCAQEMANPOHFL	1620
Qy	1561	FIANNGTVPCCPKKVVCSNTECHNGTCVNOQDASCECPFLFGGSCAQEMANPOHFL	1620
Dh	1621	GSSLVAMHGLSLPISQPMYLSLMPRTROADGVLLQAITRGRSTITTQLRBGVMSVEGT	1680
Qy	1621	GSSLVAMHGLSLPISQPMYLSLMPRTROADGVLLQAITRGRSTITTQLRBGVMSVEGT	1680
Dh	1681	GLQASSRLBEPGANQDMHQAOLAGASGPGHAILSPYGGQRBAGNLPRLHGLHLS	1740
Qy	1681	GLQASSRLBEPGANQDMHQAOLAGASGPGHAILSPYGGQRBAGNLPRLHGLHLS	1740
Dh	1741	NITVGGIPGPAGVARGFRGCLQGVAVSDTPBGSNLDPSHGHSINVBQCSLPDQDSN	1800
Qy	1741	NITVGGIPGPAGVARGFRGCLQGVAVSDTPBGSNLDPSHGHSINVBQCSLPDQDSN	1800
Dh	1801	PCPANSYCSNDMSYSCSDPGYGNCTNVCNLANFCHOSVCTRPSAPHGTYCCCPN	1860
Qy	1801	PCPANSYCSNDMSYSCSDPGYGNCTNVCNLANFCHOSVCTRPSAPHGTYCCCPN	1860
Dh	1861	YLGPFYCESTRIDQCPRGMMGHPTQCPONCDVSKGPDPCNKTSGECHKCNHTRPPGSP	1920
Qy	1861	YLGPFYCESTRIDQCPRGMMGHPTQCPONCDVSKGPDPCNKTSGECHKCNHTRPPGSP	1920
Dh	1921	CLLDCYPTGSLSRVCDPEBQCPCKRPGVIGRQCDRCNPPFAEVTNNGEVNYSQPRAI	1980
Qy	1921	CLLDCYPTGSLSRVCDPEBQCPCKRPGVIGRQCDRCNPPFAEVTNNGEVNYSQPRAI	1980
Dh	1981	EAGIMWPRTRFGLPALAPCPKSGFTAVRHCHDRHMLPNNLPNCSTITPSELKGAEBL	2040
Qy	1981	EAGIMWPRTRFGLPALAPCPKSGFTAVRHCHDRHMLPNNLPNCSTITPSELKGAEBL	2040
Dh	2041	QRNBSGLDGRSQALLLRNATORTAGYFGSPVKAAYOLAATLLAHESTORSGFSATO	2100
Qy	2041	QRNBSGLDGRSQALLLRNATORTAGYFGSPVKAAYOLAATLLAHESTORSGFSATO	2100
Dh	2101	DVHFTENLLRVGSALLDTANKRMELIQTEGGTAWLQHYEAYASALAQNMRHTYLSPE	2160
Qy	2101	DVHFTENLLRVGSALLDTANKRMELIQTEGGTAWLQHYEAYASALAQNMRHTYLSPE	2160
Dh	2161	TIYTPNIVISVVRDLKGNFAGAKLPYRYEALRGOPDLTETVILPESVRETPPVVRPAG	2220
Qy	2161	TIYTPNIVISVVRDLKGNFAGAKLPYRYEALRGOPDLTETVILPESVRETPPVVRPAG	2220
Dh	2221	PGEAQPEELARQRHRPFLSGEBAVASIITRTLAGLLPHNYDPDKSLRVPKRPDIINT	2280
Qy	2221	PGEAQPEELARQRHRPFLSGEBAVASIITRTLAGLLPHNYDPDKSLRVPKRPDIINT	2280



Db 301 BGSQSGSEFELEIDPRSGVIRTRGPVDRREEVESYQLTVEASDQGRDGPSTTAAYELSV 360  
Qy 390 EDDNDNAPORSEKKRYVQVREDVTPGA.PYLAVTASDSDKSNAYVHSHSINGNARGQFYL 449  
Db 361 EDDNDNAPORSEKKRYVQVREDVTPGA.PYLAVTASDSDKSNAYVHSHSINGNARGQFYL 420  
Qy 450 DAQGTALDVVSPLDYETTKETKYLTVRAQDGGPRPLSNVSGLVTVQVLDINDNAPIFVSTP 509  
Db 421 DAQGTALDVVSPLDYETTKETKYLTVRAQDGGPRPLSNVSGLVTVQVLDINDNAPIFVSTP 480  
Qy 510 FOATVLESVPGLTVLHVQALDADAGDNARLEFYLAVGHDPPPTINNCGWISVAEELD 569  
Db 481 FOATVLESVPGLTVLHVQALDADAGDNARLEFYLAVGHDPPPTINNCGWISVAEELD 540  
Qy 570 REEDYFSPGYEARDHGTALTAASVSVTVLDVNDNPPFTOEXYVRLNEDAAVGTSV 629  
Db 541 REEDYFSPGYEARDHGTALTAASVSVTVLDVNDNPPFTOEXYVRLNEDAAVGTSV 600  
Qy 630 VTVSAVDRDAHSVITTOITSGNTNRPSITSQSGGLVSLALPLDYKLERQYVLAATASD 689  
Db 601 VTVSAVDRDAHSVITTOITSGNTNRPSITSQSGGLVSLALPLDYKLERQYVLAATASD 660  
Qy 690 GTRDQTAQIVNVTDANTHRPVFQSHYTVVAVNEDRPAGTTVVLISATDEBTGENARITY 749  
Db 661 GTRDQTAQIVNVTDANTHRPVFQSHYTVVAVNEDRPAGTTVVLISATDEBTGENARITY 720  
Qy 750 FMEDSIFQFRIDATGAVTTOAELDYEDQVSTLTAITARDNGIPQSKDITTYLETIVNDVN 809  
Db 721 FMEDSIFQFRIDATGAVTTOAELDYEDQVSTLTAITARDNGIPQSKDITTYLETIVNDVN 780  
Qy 810 DNABQPLRDSYQGSVYEDVPPTSVLQISATDRDQSGINGRVTYFQGGDDGDGFIVEST 869  
Db 781 DNABQPLRDSYQGSVYEDVPPTSVLQISATDRDQSGINGRVTYFQGGDDGDGFIVEST 840  
Qy 870 SGYRTRRLRLDRENVAVQVTLBAVAVDKGMPARTPMETVTVLVDVNDNPPFEDDEEDVF 929  
Db 841 SGYRTRRLRLDRENVAVQVTLBAVAVDKGMPARTPMETVTVLVDVNDNPPFEDDEEDVF 900  
Qy 930 VEENSPIGLAVARVATDPDEGTNAQIMYOIVEGNIPEVQOLDIFSGELTALVOLYEDR 989  
Db 901 VEENSPIGLAVARVATDPDEGTNAQIMYOIVEGNIPEVQOLDIFSGELTALVOLYEDR 960  
Qy 990 PEYVLVITQATSAPIVSRATVAVRLLDRNDNPPVLANFELFNNTYVNRSSSPGCAIGRV 1049  
Db 961 PEYVLVITQATSAPIVSRATVAVRLLDRNDNPPVLANFELFNNTYVNRSSSPGCAIGRV 1020  
Qy 1050 PAHDPOISDSITVSPERGENELSTVLINASTGELKLSRALDNNRPLEAIMSTLVSDGVHSV 1109  
Db 1021 PAHDPOISDSITVSPERGENELSTVLINASTGELKLSRALDNNRPLEAIMSTLVSDGVHSV 1080  
Qy 1110 TAQCALRVITITDEMILTHSITLRLSDMSPERFLSPILGLFIQAVAAATLAPPDVVVFN 1169  
Db 1081 TAQCALRVITITDEMILTHSITLRLSDMSPERFLSPILGLFIQAVAAATLAPPDVVVFN 1140  
Qy 1170 QRDTPADPGHILINSLVSGVPPGPGGPPPLPSEDLQERLYINRSLTALISAQVLPFD 1229  
Db 1141 QRDTPADPGHILINSLVSGVPPGPGGPPPLPSEDLQERLYINRSLTALISAQVLPFD 1200  
Qy 1230 NICIREPENMRCSVLRPDSAPFIASSSVLRPHHPVGLACRCGPGTGYCEREV 1289  
Db 1201 NICIREPENMRCSVLRPDSAPFIASSSVLRPHHPVGLACRCGPGTGYCEREV 1260  
Qy 1290 DLCTSRPCGPHGRCSRREGYTCLCRDGYTGEHCEVSARSRCITPGVCKNGCTCVNLLVG 1349  
Db 1261 DLCTSRPCGPHGRCSRREGYTCLCRDGYTGEHCEVSARSRCITPGVCKNGCTCVNLLVG 1320  
Qy 1350 GPKCDPCSGDEBKPYCOVTTSSFPANSTITTRGLRQRFHFTLASFATKEDGLLLNGR 1409  
Db 1321 GPKCDPCSGDEBKPYCOVTTSSFPANSTITTRGLRQRFHFTLASFATKEDGLLLNGR 1380  
Qy 1410 FNEKHDPALEVIQEOVOLTSAGSSTTVSPVVGVSDDQWHTVOLKYNKPLLGOTG 1469

Db 1381 FNEKHDPALEVIQEOVOLTSAGSSTTVSPVVGVSDDQWHTVOLKYNKPLLGOTG 1440  
Qy 1470 LPQSPBOKVAVTVYDGDCTGVALRFGSVLGNVSCAAOGTQGGSKSLDTGPLLGGVP 1529  
Db 1441 LPQSPBOKVAVTVYDGDCTGVALRFGSVLGNVSCAAOGTQGGSKSLDTGPLLGGVP 1500  
Qy 1530 DLBPSPVNRQOPVGCNRMLQVDSRHI DNADFIANNGTVPGCBAKXNCDNSNCHNGGTC 1589  
Db 1501 DLBPSPVNRQOPVGCNRMLQVDSRHI DNADFIANNGTVPGCBAKXNCDNSNCHNGGTC 1560  
Qy 1590 VNOQWDAFSCBCEPLGREGKSCAQMENAPQHFQSSLVAMHGLSLPIQPMWLSIMFTTQRA 1649  
Db 1561 VNOQWDAFSCBCEPLGREGKSCAQMENAPQHFQSSLVAMHGLSLPIQPMWLSIMFTTQRA 1620  
Qy 1650 DGVILQAITRGRSTITLQIREGHVMSVGTGLOASSLRLEPBRANDGMVHQAQLLQAS 1709  
Db 1621 DGVILQAITRGRSTITLQIREGHVMSVGTGLOASSLRLEPBRANDGMVHQAQLLQAS 1680  
Qy 1710 GGFPAHILSPDYQOQABGNLGRHLGLSNITVGGIPQAPAGVARGRGCLQGVAVSD 1769  
Db 1681 GGFPAHILSPDYQOQABGNLGRHLGLSNITVGGIPQAPAGVARGRGCLQGVAVSD 1740  
Qy 1770 TPEGVNSLDPSSHGHSINVEQGSCLPDPDPSNPPCPANSYCSNDMDYSJCSGDPGYDNDCT 1829  
Db 1741 TPEGVNSLDPSSHGHSINVEQGSCLPDPDPSNPPCPANSYCSNDMDYSJCSGDPGYDNDCT 1800  
Qy 1830 NVCDLNPCEHQSCTKRPAPAGYTCECPNLYGPFCEFRIDOPCPRGMMGHPTGCPNC 1889  
Db 1801 NVCDLNPCEHQSCTKRPAPAGYTCECPNLYGPFCEFRIDOPCPRGMMGHPTGCPNC 1860  
Qy 1890 DVSQKDFDPCNKTSGECHCKENHRRPGSPPTCLCDCTPGSLSRVCDPEDQCPCKPGV 1949  
Db 1861 DVSQKDFDPCNKTSGECHCKENHRRPGSPPTCLCDCTPGSLSRVCDPEDQCPCKPGV 1920  
Qy 1950 IGRQCRCONPFAEYVTNGCEVYVYDSCPRATBAGIMWPPTRFGLPAAAPCPKSPGTAVR 2009  
Db 1921 IGRQCRCONPFAEYVTNGCEVYVYDSCPRATBAGIMWPPTRFGLPAAAPCPKSPGTAVR 1980  
Qy 2010 HCDEHGMPLPMLFNCTSIITFSELKGFARLQBNESGLDGRSQOULALLRNATQHTAGY 2069  
Db 1981 HCDEHGMPLPMLFNCTSIITFSELKGFARLQBNESGLDGRSQOULALLRNATQHTAGY 2040  
Qy 2070 FGSQVAVAYOLATRLAHSTQRFGLSATQDVHFTBNILRVGSALLDTRANKHMLIQ 2129  
Db 2041 FGSQVAVAYOLATRLAHSTQRFGLSATQDVHFTBNILRVGSALLDTRANKHMLIQ 2100  
Qy 2130 TBGSTMLOHRYAAYASALANONRHITYLSPFTVTEINIYISVRLDKGNPAGAKLPRYEA 2189  
Db 2101 TBGSTMLOHRYAAYASALANONRHITYLSPFTVTEINIYISVRLDKGNPAGAKLPRYEA 2160  
Qy 2190 LRGSOPPDLETITVILPESVFERETPPVVRPAGPBAQEPBELARORRHPELSQGEAVASY 2249  
Db 2161 LRGSOPPDLETITVILPESVFERETPPVVRPAGPBAQEPBELARORRHPELSQGEAVASY 2220  
Qy 2250 IYRTTLAAGLPHNYDDPKSLRVPKRPINTPVASISVHDEBELLRALDKVTVYQFRLL 2309  
Db 2221 IYRTTLAAGLPHNYDDPKSLRVPKRPINTPVASISVHDEBELLRALDKVTVYQFRLL 2280  
Qy 2310 ETEERTKPICVFNHSHILVSGTGWMSARGCEVFRSHSHVSCQNMHTSPAVLMDVSRRE 2369  
Db 2281 ETEERTKPICVFNHSHILVSGTGWMSARGCEVFRSHSHVSCQNMHTSPAVLMDVSRRE 2240  
Qy 2370 NGBILPDKLTLYVALGVTLAALLTFFPFLTLRLISNQHIGIRNLTALGLAQLVFLG 2429  
Db 2341 NGBILPDKLTLYVALGVTLAALLTFFPFLTLRLISNQHIGIRNLTALGLAQLVFLG 2400  
Qy 2430 INQADLPRACTVAILLHFLYLCFSPALLBAHLTRALTEVDVNTGPMRFYTMGMV 2489  
Db 2401 INQADLPRACTVAILLHFLYLCFSPALLBAHLTRALTEVDVNTGPMRFYTMGMV 2460  
Qy 2490 PAFITGLAVGLDEPGGNPFCWLSIYDTLINSPAGVAPAVMSVFLYTLARASCAQ 2549  
Db 2461 PAFITGLAVGLDEPGGNPFCWLSIYDTLINSPAGVAPAVMSVFLYTLARASCAQ 2520



QY 2550 RQFEKKGPVSGIQPSFVALLLSATWLLALSVNSDTLLFHYLPATCNCTIGPPIFLSY 2609  
DB 2521 RQFEKKGPVSGIQPSFVALLLSATWLLALSVNSDTLLFHYLPATCNCTIGPPIFLSY 2580  
QY 2610 VVLSKEVRKALKACSKRPDPALTTKSTLTSYNCPSPYADGRLYOPYGDSSAGSLHST 2669  
DB 2581 VVLSKEVRKALKACSKRPDPALTTKSTLTSYNCPSPYADGRLYOPYGDSSAGSLHST 2640  
QY 2670 SRSGKQPSYIPLLEBSALNPGQPPGLGDPGSLFLEGQDQHPDPTDSDLSLEDD 2729  
DB 2641 SRSGKQPSYIPLLEBSALNPGQPPGLGDPGSLFLEGQDQHPDPTDSDLSLEDD 2700  
QY 2730 QSGSYASTHSSDBEEBEEBEEBEEBAPFEGCGMSLIGPGEFRLPHSTPDGSGPGKAP 2789  
DB 2701 QSGSYASTHSSDBEEBEEBEEBEEBAPFEGCGMSLIGPGEFRLPHSTPDGSGPGKAP 2760  
QY 2790 WPGDFGTAKSSGNGAPBERLRENGDALREGSLGPLPSSSQPHKGLTKKCLPTIS 2849  
DB 2761 WPGDFGTAKSSGNGAPBERLRENGDALREGSLGPLPSSSQPHKGLTKKCLPTIS 2820  
QY 2850 KSSLRLPLEQCTGSSRSGSASRSGRGPPEPPRQSLQEQNLGWPFIAMSIIKAGTVDE 2909  
DB 2821 KSSLRLPLEQCTGSSRSGSASRSGRGPPEPPRQSLQEQNLGWPFIAMSIIKAGTVDE 2880  
QY 2910 DSGSGFLLPNNFLH 2923  
DB 2881 DSGSGFLLPNNFLH 2894

RESULT 5  
AAU07053  
ID AAU07053 standard, Protein, 2956 AA.  
AC AAU07053;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human Flamingo polypeptide.  
XX  
Flamingo; human; G-protein coupled receptor; diabetes; protozoan; cancer;  
signal transduction pathway; bacterial; fungal; viral; anorexia; bulimia;  
asthma; Parkinson's disease; acute heart failure; HIV; obesity; anxiety;  
hypotension; hypertension; urinary retention; angina pectoris; allergy;  
myocardial infarction; stroke; ulcer; schizophrenia; psychotic disorder;  
benign prostatic hyperplasia; neurological disorder; manic depression;  
delirium; dementia; severe mental retardation; Huntington's disease;  
Gilles de la Tourette's syndrome; antibacterial; antifungal; antiviral;  
antiprotozoal; anti-HIV; anorectic; antianorectic; antiparkinsonian;  
cardiac; cerebroprotective; neuroprotective; antidepressant;  
anticonvulsant; antitense therapy; gene therapy.  
XX  
Homo sapiens.  
XX  
WO200161003-A1.  
PN  
XX  
PD 23-AUG-2001.  
XX  
PF 19-FEB-2001; 2001WO-GB00680.  
XX  
PR 19-FEB-2000; 2000GB-0004196.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Testa TT;  
XX  
DR WPI: 2001-502792/55.  
XX  
N-PSDB; AAS11677.  
PT An isolated Flamingo polypeptide useful for treating diseases such as  
PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
XX osteoporosis -

PS Claim 2; Page 28-29; 66pp; English.  
XX  
CC The sequence represents a human Flamingo polypeptide. Flamingo is a  
CC member of the G-protein coupled receptor family, which is involved in  
CC signal transduction pathways. By screening to identify compounds that  
CC stimulate or inhibit the function or level of the protein, treatments can  
CC be developed for various diseases and bacterial, fungal, protozoan and  
CC viral infections, including HIV, cancer, diabetes, obesity, anorexia,  
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, angina pectoris, myocardial infarction,  
CC stroke, ulcers, allergies and benign prostatic hyperplasia. Also  
CC treatable are psychotic and neurological disorders such as anxiety,  
CC schizophrenia, manic depression, delirium, dementia, severe mental  
CC retardation, Huntington's disease and Gilles de la Tourette's syndrome.  
XX  
SQ Sequence 2956 AA;

Query Match 97.1%; Score 2837; DB 22; Length 2956;  
Best Local Similarity 100.0%; Fred. No. 0;  
Matches 2837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESPATGVLPPPPPLILLLLPPPLGDPVPCRSISGRSGSACAPMGWLCPS 60  
DB 1 MESPATGVLPPPPPLILLLLPPPLGDPVPCRSISGRSGSACAPMGWLCPS 60  
QY 61 SASNIMLYTSRCRDAGTELGHVPHHDGLRVWCPSSEAHIPLPAPBECGPMSCRLGIG 120  
DB 61 SASNIMLYTSRCRDAGTELGHVPHHDGLRVWCPSSEAHIPLPAPBECGPMSCRLGIG 120  
QY 121 GHLSFGKLTLPBEPCLKAPRLRCQSCGLAAPGIRABRRBPBESLGGRRRNVTAAQ 180  
DB 121 GHLSFGKLTLPBEPCLKAPRLRCQSCGLAAPGIRABRRBPBESLGGRRRNVTAAQ 180  
QY 181 FQPPSYQATVPENQAGTFVASISRAIDPBGEGAGRLLEYTMDALPDSRSQFSLDPVTA 240  
DB 181 FQPPSYQATVPENQAGTFVASISRAIDPBGEGAGRLLEYTMDALPDSRSQFSLDPVTA 240  
QY 241 VTTAEELDEREYKSTVFRVTAQDHGMPRRSALATLTITVTNDHDVFEQGEYKESLRE 300  
DB 241 VTTAEELDEREYKSTVFRVTAQDHGMPRRSALATLTITVTNDHDVFEQGEYKESLRE 300  
QY 301 NIEVGYEVLTVATATGDAAPPNNILYRLLEGSGSPSEVFEIDPDSGVRTGPVREEV 360  
DB 301 NIEVGYEVLTVATATGDAAPPNNILYRLLEGSGSGSPSEVFEIDPDSGVRTGPVREEV 360  
QY 361 ESYQUTVEASDQGRDPSPSTTAAVFLSVEDNDNAPQPSERKRYVQVREEDVTPGAPVLR 420  
DB 361 ESYQUTVEASDQGRDPSPSTTAAVFLSVEDNDNAPQPSERKRYVQVREEDVTPGAPVLR 420  
QY 421 VITASDRDKGSNAVVHYSIMSGNARQFTYLDAGTALDVVSPLDYETTKETTLRVAQDDG 480  
DB 421 VITASDRDKGSNAVVHYSIMSGNARQFTYLDAGTALDVVSPLDYETTKETTLRVAQDDG 480  
QY 481 RPLSNVSGLVTVQVLDINDNAPFVSTPPOATVLESPVGLTALHVOALDADAGNML 540  
DB 481 RPLSNVSGLVTVQVLDINDNAPFVSTPPOATVLESPVGLTALHVOALDADAGNML 540  
QY 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVEDFYSPGVABRDGTPLATASASVTV 600  
DB 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVEDFYSPGVABRDGTPLATASASVTV 600  
QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYQTISGNTRNPSITS 660  
DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYQTISGNTRNPSITS 660  
QY 661 OSGGGLVSLALPLDYKLEROYVLAVTASDGTROPDAQIVVNTDANTHRPVSQSHYTIN 720  
DB 661 OSGGGLVSLALPLDYKLEROYVLAVTASDGTROPDAQIVVNTDANTHRPVSQSHYTIN 720  
QY 721 VNEBRPAGTIVVLSATBDETCENARITYFMEDSIPQFRIDADTAVTTOAELDYEDQVS 780  
DB 721 VNEBRPAGTIVVLSATBDETCENARITYFMEDSIPQFRIDADTAVTTOAELDYEDQVS 780

QY 781 YTLAITARDNGI POKSOTTYLEIIVNDVNDNAPOLFARDSYGSGVYEDVPEPTSYLQISAT 840  
Db 781 YTLAITARDNGI POKSOTTYLEIIVNDVNDNAPOLFARDSYGSGVYEDVPEPTSYLQISAT 840  
QY 841 DRDGLNGRVVYTYOGGDDGDDDFIVESTSGIVARTLRLEBENNAQVYLAAYDKMPP 900  
Db 841 DRDGLNGRVVYTYOGGDDGDDDFIVESTSGIVARTLRLEBENNAQVYLAAYDKMPP 900  
QY 901 ARTEMEVTVTVLDVNDNPPVEODEFDVFEVENSPIGLAVARTATPDDEGTNAQIMYQI 960  
Db 901 ARTEMEVTVTVLDVNDNPPVEODEFDVFEVENSPIGLAVARTATPDDEGTNAQIMYQI 960  
QY 961 VEGNIPEVFOQDIDSGEITLALVDLDYEDRPEYVLYIQATSAPLYSRATVHRLIDRDNDP 1020  
Db 961 VEGNIPEVFOQDIDSGEITLALVDLDYEDRPEYVLYIQATSAPLYSRATVHRLIDRDNDP 1020  
QY 1021 PVLENPELIFNNYNTNNSSPFGAIGRPVPAHDIDISLTYSPERGENELSLVILNASTG 1080  
Db 1021 PVLENPELIFNNYNTNNSSPFGAIGRPVPAHDIDISLTYSPERGENELSLVILNASTG 1080  
QY 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTACALRVTLITDEMILTHSITLRLEDMSPER 1140  
Db 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTACALRVTLITDEMILTHSITLRLEDMSPER 1140  
QY 1141 FLSPILGLFIOAVATATLTPPDHVVVNVQORDTAPGGHILNVSLSVQPPGCGPPFL 1200  
Db 1141 FLSPILGLFIOAVATATLTPPDHVVVNVQORDTAPGGHILNVSLSVQPPGCGPPFL 1200  
QY 1201 PSEBLOBLVYNRSLTALISAQVLPDPDNLCLREPCENYRCVSLRPFSSAPFIASSS 1260  
Db 1201 PSEBLOBLVYNRSLTALISAQVLPDPDNLCLREPCENYRCVSLRPFSSAPFIASSS 1260  
QY 1261 VLFPRHIVGVGLRCRCPGFTGDCETEVDLCYSRPGCGPHRCRSREGYTCICRDSYTG 1320  
Db 1261 VLFPRHIVGVGLRCRCPGFTGDCETEVDLCYSRPGCGPHRCRSREGYTCICRDSYTG 1320  
QY 1321 EHCVSARSRGCTGCKNGCTCVNLVYGKCCPSGDPEKPYCOYTTRSPFAHSITTF 1380  
Db 1321 EHCVSARSRGCTGCKNGCTCVNLVYGKCCPSGDPEKPYCOYTTRSPFAHSITTF 1380  
QY 1381 RGLRQRHFPTLAFATKERDGLLNYGRFNEKHDPALEVIQOVOLTSRAGSITTVS 1440  
Db 1381 RGLRQRHFPTLAFATKERDGLLNYGRFNEKHDPALEVIQOVOLTSRAGSITTVS 1440  
QY 1441 PVPFGVSDGQWHTVOLKTYNKPILLGOTGLPQGSBQKVAVVTDGCDTGVALRFGSVLG 1500  
Db 1441 PVPFGVSDGQWHTVOLKTYNKPILLGOTGLPQGSBQKVAVVTDGCDTGVALRFGSVLG 1500  
QY 1501 NYSQAAGTQGSKSLDLTGPILLGVPDLPSFPVPRMRQFVGCMBNLQVDSRHIMAD 1560  
Db 1501 NYSQAAGTQGSKSLDLTGPILLGVPDLPSFPVPRMRQFVGCMBNLQVDSRHIMAD 1560  
QY 1561 FIANGTVPGCPAKGNCDSVTGANGGTCVQNMOPAFSCPCPLGFGKSCAQBMANPQFL 1620  
Db 1561 FIANGTVPGCPAKGNCDSVTGANGGTCVQNMOPAFSCPCPLGFGKSCAQBMANPQFL 1620  
QY 1621 GSSIVAHGSLSPISQPMYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVLVEGT 1680  
Db 1621 GSSIVAHGSLSPISQPMYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVLVEGT 1680  
QY 1681 GLQASSRLRBRGRANDDMHQAOLALASGSGRAIISFDYGOORAAGNIGPRLHGHLS 1740  
Db 1681 GLQASSRLRBRGRANDDMHQAOLALASGSGRAIISFDYGOORAAGNIGPRLHGHLS 1740  
QY 1741 NITVGGIPGAGVARGFRCLOGVRSVDPREGVNSLDPSSHGSIINTEGCSLPDPCSN 1800  
Db 1741 NITVGGIPGAGVARGFRCLOGVRSVDPREGVNSLDPSSHGSIINTEGCSLPDPCSN 1800  
QY 1801 PCPANSYCSNDWDSYSCDPRGYTGDNCTNVCIDINPCCHOSVCTRKSAPHGTYCPCPN 1860  
Db 1801 PCPANSYCSNDWDSYSCDPRGYTGDNCTNVCIDINPCCHOSVCTRKSAPHGTYCPCPN 1860  
QY 1861 YLGPYCEFRIDQPCPRGMWHPITCGPCNCDVSKGFPDQCKMTSGECHKCNHTRPPSPPT 1920

Db 1861 YLGPYCEFRIDQPCPRGMWHPITCGPCNCDVSKGFPDQCKMTSGECHKCNHTRPPSPPT 1920  
QY 1921 CILCDYPPGSGSRVCDPEDGOCPCRPVYGRCCDCDNFPAVTNNGCVNDSQBRAL 1980  
Db 1921 CILCDYPPGSGSRVCDPEDGOCPCRPVYGRCCDCDNFPAVTNNGCVNDSQBRAL 1980  
QY 1981 EAGIWPRTFGLPAAPCPKSGFATAVRCHDRMCLPNNFNCTSIYESLKGFAERL 2040  
Db 1981 EAGIWPRTFGLPAAPCPKSGFATAVRCHDRMCLPNNFNCTSIYESLKGFAERL 2040  
QY 2041 QRNESGLDSGRSQQLALRNATQHTAGIFGSDVAVAYQATLTLAHESTORGEGSATQ 2100  
Db 2041 QRNESGLDSGRSQQLALRNATQHTAGIFGSDVAVAYQATLTLAHESTORGEGSATQ 2100  
QY 2101 DVAFTEMLRVSGALLDPTANKRMELIQQTEGTALMLQHYEAYASALQNMHTYLSPF 2160  
Db 2101 DVAFTEMLRVSGALLDPTANKRMELIQQTEGTALMLQHYEAYASALQNMHTYLSPF 2160  
QY 2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGEPDLETTVILPESVPERETPPVVRPAG 2220  
Db 2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGEPDLETTVILPESVPERETPPVVRPAG 2220  
QY 2221 PGSAQBPBEELARRQRRHPELSQGEAVASYITRTLAGLLPHNYDPDKRSILVPEKPIINT 2280  
Db 2221 PGSAQBPBEELARRQRRHPELSQGEAVASYITRTLAGLLPHNYDPDKRSILVPEKPIINT 2280  
QY 2281 PVVISVHDEBELPALDKPVTVOQRLLETERTPICVFMHSLTVSGTQMSARGCE 2340  
Db 2281 PVVISVHDEBELPALDKPVTVOQRLLETERTPICVFMHSLTVSGTQMSARGCE 2340  
QY 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGYTLAALLTFPEFLT 2400  
Db 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGYTLAALLTFPEFLT 2400  
QY 2401 LRILRNGGIRRNLTALAGLQVFLGIGNADLPACTIVAILHPLYLCTFSWALLR 2460  
Db 2401 LRILRNGGIRRNLTALAGLQVFLGIGNADLPACTIVAILHPLYLCTFSWALLR 2460  
QY 2461 ALHLRYALTEVRDVTNTPMRFTYMLGMGPAPITGLAVGLDPGYGNPFCMLIYDTLL 2520  
Db 2461 ALHLRYALTEVRDVTNTPMRFTYMLGMGPAPITGLAVGLDPGYGNPFCMLIYDTLL 2520  
QY 2521 WSPAGVAVAVSVFLYTLAARASCAQORQGEKKGPVSGLOPSFAVLLLSATVLLAL 2580  
Db 2521 WSPAGVAVAVSVFLYTLAARASCAQORQGEKKGPVSGLOPSFAVLLLSATVLLAL 2580  
QY 2581 LSVNSDTLLEHYLFATCNCIOGPFITLSTVYLSKEVRKALKACSRKPSBPPALTTKSTL 2640  
Db 2581 LSVNSDTLLEHYLFATCNCIOGPFITLSTVYLSKEVRKALKACSRKPSBPPALTTKSTL 2640  
QY 2641 TSSYNCPSPYADRLTQPYGDSAGSLHSTRSGKSQPSYIFLLAREESALNPGQGPGLG 2700  
Db 2641 TSSYNCPSPYADRLTQPYGDSAGSLHSTRSGKSQPSYIFLLAREESALNPGQGPGLG 2700  
QY 2701 DPGSLFLEGOQDQHDPTDSDLSLEDDQSGSYASTSSDSEEBEERBEAFAFGEOG 2760  
Db 2701 DPGSLFLEGOQDQHDPTDSDLSLEDDQSGSYASTSSDSEEBEERBEAFAFGEOG 2760  
QY 2761 WDSLAPGAEURLPHSTPDKDGGPGKAPWPDPGTAAESSGNGAPREXRIRNGALSR 2820  
Db 2761 WDSLAPGAEURLPHSTPDKDGGPGKAPWPDPGTAAESSGNGAPREXRIRNGALSR 2820  
QY 2821 EGSLGPLPSSAQPHKG 2837  
Db 2821 EGSLGPLPSSAQPHKG 2837  
RESULT 6  
AAU74826  
ID AAU74826 standard; protein; 2936 AA.  
XX  
AC AAU74826;

XX 23-APR-2002 (first entry)  
 XX Human REPTR 9 protein.  
 DE  
 XX  
 KM REPTR: human; antiinflammatory; cytostatic; immunosuppressive;  
 KM antiviral; anti-HIV; antitumor; anticonvulsant; nootropic;  
 KM neuroprotective; anti-allergic; antibody; immunogen; endometriosis;  
 KM gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KM Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KM autoimmune disorder; hypothyroidism; Kallman's disease;  
 KM endocrine disease; inflammatory disease; infertility; receptor;  
 KM acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KM allergic; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KM systemic lupus erythematosus; cell proliferative disorder;  
 KM cancer; developmental disorder; Duchenne muscular dystrophy;  
 KM Becker muscular dystrophy; neurological disorder; epilepsy;  
 KM Alzheimer's disease; Huntington's disease; reproductive disorder.  
 XX  
 OS Homo sapiens.  
 PN  
 XX  
 PN W0200198354-A2.  
 PD  
 XX  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-US19942.  
 XX  
 XX 21-JUN-2000; 2000US-214027P. — *Chick*  
 XX 25-JUN-2000; 2000US-228045P.  
 XX 12-DEC-2000; 2000US-255104P.  
 XX  
 PA (INCY- ) INCYTE GENOMICS INC.  
 PI Griffen JA, Kalliock DA, Tribouley CM, Yue H, Nguyen DB, Tang YT;  
 PI Lal P, Policky JL, Azimzai Y, Lu DM, Grail R, Yao MG, Burford N;  
 PI Hafajia AJA, Baughn MR, Bandman O, Patterson C, Yang J, Xu Y;  
 PI Gandhi AR, Warren BA, Ding L, Sanjanwala NS, Duggan BM,  
 XX  
 DR WPI; 2002-090432/12.  
 XX N-PSDB; ABK15177.  
 XX  
 PT Twelve human receptors (referred to as REPTR-1 to REPTR-12), useful in  
 PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
 PT proliferative (e.g. cancer) disorders -  
 XX  
 PS Claim 53; Page 131-138; 157pp; English.  
 XX  
 CC This invention relates to twelve human receptors cDNA sequences  
 CC referred to as REPTR-1 to REPTR-12), and the proteins encoded thereby.  
 CC The proteins of the invention may have antiinflammatory, cytostatic,  
 CC immunosuppressive, antiviral, anti-HIV, antitumor, muscular active  
 CC general, anticonvulsant, nootropic, neuroprotective, antiallergic  
 CC activities. The sequences of the invention may be used to produce REPTR  
 CC agonists or antagonists, and the protein sequences may be used to raise  
 CC anti-REPTR antibodies. These molecules and the REPTR polynucleotides and  
 CC polypeptides of the invention are useful in the diagnosis, treatment and  
 CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,  
 CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),  
 CC endocrine (e.g. hypothyroidism disorder, Kallman's disease), autoimmune/  
 CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS)),  
 CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,  
 CC multiple sclerosis, systemic lupus erythematosus), cell proliferative  
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
 CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,  
 CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)  
 CC disorders. Numerous other examples of each disorder are given in the  
 CC specification. The present sequence represents the human REPTR9 protein  
 CC sequence of the invention.  
 XX  
 XX Sequence 2936 AA;  
 Query Match 67.0%; Score 1958; DB 23; Length 2936;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1958; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 PPPPLLLLLLLLLPPPLGDDGVGPPCSLGRSGSGAACPWCMCLPPSSASNNLMTYSNC 72  
 DB 14 PPPPLLLLLLLLLPPPLGDDGVGPPCSLGRSGSGAACPWCMCLPPSSASNNLMTYSNC 73  
 QY 73 RDAGELTGLVPHHDLGRLVWCPESEBAHPLPAPAGCPCWSCRLGIGHLSPQGLTLP 132  
 DB 74 RDAGELTGLVPHHDLGRLVWCPESEBAHPLPAPAGCPCWSCRLGIGHLSPQGLTLP 133  
 QY 133 EHPPLCAKRLRCQSCCKLAQAPGLRAGERSPEESLGGRRKRVNTAPQCPESYQATVP 192  
 DB 134 EHPPLCAKRLRCQSCCKLAQAPGLRAGERSPEESLGGRRKRVNTAPQCPESYQATVP 193  
 QY 193 NOPAGPVASLAIIDDEGRARLRYTMALPDSRNOFPISLDPVGAATTAEELEDRNK 252  
 DB 194 NOPAGPVASLAIIDDEGRARLRYTMALPDSRNOFPISLDPVGAATTAEELEDRNK 253  
 QY 253 STHVFRTAODHGMPPRSALATLTLLVTDNDHPVFEQOEYKESLRENLVGEVLTVR 312  
 DB 254 STHVFRTAODHGMPPRSALATLTLLVTDNDHPVFEQOEYKESLRENLVGEVLTVR 313  
 QY 313 ATDGAAPVNAITLRLBSGSGSPSEVFEDIPRSGVITRGPVDRREVESYQLTVASDQ 372  
 DB 314 ATDGAAPVNAITLRLBSGSGSPSEVFEDIPRSGVITRGPVDRREVESYQLTVASDQ 373  
 QY 373 GRDGRSTTAAPVLSVEDNDNNAPOFSERKRVVQVREBVTGAPLRTATSDRQGSNA 432  
 DB 374 GRDGRSTTAAPVLSVEDNDNNAPOFSERKRVVQVREBVTGAPLRTATSDRQGSNA 433  
 QY 433 VVHYSIMSGNARGOFLDQALDVSPLDYETTKETTLRYAODGGRPLSMVSGLV 492  
 DB 434 VVHYSIMSGNARGOFLDQALDVSPLDYETTKETTLRYAODGGRPLSMVSGLV 493  
 QY 493 VQVLDINDNAPLFTSTPQATVLESYPLGLVLAHQALDADAGNARLERLAGVGHDP 552  
 DB 494 VQVLDINDNAPLFTSTPQATVLESYPLGLVLAHQALDADAGNARLERLAGVGHDP 553  
 QY 553 FTINNGTGMISVAELDEEVDVPSYGVARHGPALFASASVSTVLDVNDNNTFPQ 612  
 DB 554 FTINNGTGMISVAELDEEVDVPSYGVARHGPALFASASVSTVLDVNDNNTFPQ 613  
 QY 613 PEYVLNEDAAVGVSVTVASVDRDAHSVITYQITSGTNRNRFSTSGSGGLVSLAP 672  
 DB 614 PEYVLNEDAAVGVSVTVASVDRDAHSVITYQITSGTNRNRFSTSGSGGLVSLAP 673  
 QY 673 LDYKLRQVYLAVTASDGTROTQIIVNVTANTHRPVFGSHYTVVNEDEPACTTV 732  
 DB 674 LDYKLRQVYLAVTASDGTROTQIIVNVTANTHRPVFGSHYTVVNEDEPACTTV 733  
 QY 733 LISATDEDEGENARITTFMEDSIPOFRIDADGAVTQALDYEDOVSTLTATANDGI 792  
 DB 734 LISATDEDEGENARITTFMEDSIPOFRIDADGAVTQALDYEDOVSTLTATANDGI 793  
 QY 793 POKSDTTLLEIIVNDVNDNAPQLRDSYGSYVEDVPPTSYQISATDRDGLNGRV 852  
 DB 794 POKSDTTLLEIIVNDVNDNAPQLRDSYGSYVEDVPPTSYQISATDRDGLNGRV 853  
 QY 853 TTQGGDDGCGDITVESTGIVTTLRLDRBNVAQVYLAAYVDKMPARTPMEVTVTL 912  
 DB 854 TTQGGDDGCGDITVESTGIVTTLRLDRBNVAQVYLAAYVDKMPARTPMEVTVTL 913  
 QY 913 DVNDNPVEQGEFVDFVENSPIGLAVARVATPDDEGTNAQIMQIVEGNIPEVFO 972  
 DB 914 DVNDNPVEQGEFVDFVENSPIGLAVARVATPDDEGTNAQIMQIVEGNIPEVFO 973  
 QY 973 IFSGELTALVDLDYDRPEYVLVIQATSAPLVSRATVHRLLDNDNPEVLGNFELFN 1032  
 DB 974 IFSGELTALVDLDYDRPEYVLVIQATSAPLVSRATVHRLLDNDNPEVLGNFELFN 1033  
 QY 1033 YVTNRSSFPGCAIGRVPAHDPIDSLTYSERGENLSVLINASTGLKLSRALDNR 1092

Db 1034 VYNNRSSFGPGAGIARVPAHDPDISDLSLTFFERGENELSLVLNASTGELTLSPALDNMR 1093  
 QY 1093 PLEAIMSVLVSDGVSYATAOCALRTVITITDEMTHTSITLRLSDMSPERFLSPILGLFIQA 1152  
 Db 1094 PLEAIMSVLVSDGVSYATAOCALRTVITITDEMTHTSITLRLSDMSPERFLSPILGLFIQA 1153  
 QY 1153 VAATLATPPDHVVVFNVRDTPADPGHILNLSVGVQPPGPGGPPFLPSEDLQERLYTN 1212  
 Db 1154 VAATLATPPDHVVVFNVRDTPADPGHILNLSVGVQPPGPGGPPFLPSEDLQERLYTN 1213  
 QY 1213 RSLTALISAQVLPDPDDNICLREPCENMRCSVSLRFPSSAPFLASSSVLRPPIHPVGL 1272  
 Db 1214 RSLTALISAQVLPDPDDNICLREPCENMRCSVSLRFPSSAPFLASSSVLRPPIHPVGL 1273  
 QY 1273 RCRCPPGFTGDCYCEFEVDLCYSPFCGPRGCRSREGGYTCCRDGYTSEHCVSARSRC 1332  
 Db 1274 RCRCPPGFTGDCYCEFEVDLCYSPFCGPRGCRSREGGYTCCRDGYTSEHCVSARSRC 1333  
 QY 1333 TPGVCNKGTCVNLIVGFKCDPCSGDFEKPQCVTTTSPPAHSFTFRGLRQRFHTLA 1392  
 Db 1334 TPGVCNKGTCVNLIVGFKCDPCSGDFEKPQCVTTTSPPAHSFTFRGLRQRFHTLA 1393  
 QY 1393 LSPATKREDDGLLTVNGRPNRKNDFALAVIOBOVLTSAGESTTVSPVPVGGVSDQW 1452  
 Db 1394 LSPATKREDDGLLTVNGRPNRKNDFALAVIOBOVLTSAGESTTVSPVPVGGVSDQW 1453  
 QY 1453 HTVOLKTYNKPDLAQGTGLPQSPSEBOKVAVVTVDGCDTGVALRFGSVLGNYSCAAQGTGG 1512  
 Db 1454 HTVOLKTYNKPDLAQGTGLPQSPSEBOKVAVVTVDGCDTGVALRFGSVLGNYSCAAQGTGG 1513  
 QY 1513 SKKSLDLTGPIILGQVPLPSSFPVRMRQFVGCNENI QVDSRHIMADPIANNGTVPQCP 1572  
 Db 1514 SKKSLDLTGPIILGQVPLPSSFPVRMRQFVGCNENI QVDSRHIMADPIANNGTVPQCP 1573  
 QY 1573 AKKQVCSNTGNGCTCYNOWDABSCCEPLRGGSCQOEBANPQHPFGSSLVAMHGSL 1632  
 Db 1574 AKKQVCSNTGNGCTCYNOWDABSCCEPLRGGSCQOEBANPQHPFGSSLVAMHGSL 1633  
 QY 1633 PISQPMWYLSMFRTRQADGVLLQAITRGRSTTTLQLRGHVWLVEGTGLQASSLRLEPG 1692  
 Db 1634 PISQPMWYLSMFRTRQADGVLLQAITRGRSTTTLQLRGHVWLVEGTGLQASSLRLEPG 1693  
 QY 1693 RANDGDMHAOLALGASGPGHAILSPYGGQRAAGNTGPRILHGLHLSNTVGGIPGAG 1752  
 Db 1694 RANDGDMHAOLALGASGPGHAILSPYGGQRAAGNTGPRILHGLHLSNTVGGIPGAG 1753  
 QY 1753 GVANGFRGCTLGVRVSDTPBEGVNSLDPBHGHSINVEGCSLPDPCDSNPPCANSTCSNDW 1812  
 Db 1754 GVANGFRGCTLGVRVSDTPBEGVNSLDPBHGHSINVEGCSLPDPCDSNPPCANSTCSNDW 1813  
 QY 1813 DSYSCSDPGYGGNCTVNCDLNPEHOSVCTRKPASPHGYTCECPYNYLGPYCETRIQ 1872  
 Db 1814 DSYSCSDPGYGGNCTVNCDLNPEHOSVCTRKPASPHGYTCECPYNYLGPYCETRIQ 1873  
 QY 1873 PCPRGMWHPPTGCPNCNVSXKGFDPDKNTSGECHKENHYPGSPPTCLLDCYPTGSL 1932  
 Db 1874 PCPRGMWHPPTGCPNCNVSXKGFDPDKNTSGECHKENHYPGSPPTCLLDCYPTGSL 1933  
 QY 1933 SRVCDPBDGQCPCKRGVIGRCODRCNDNPFAYTNGCE 1970  
 Db 1934 SRVCDPBDGQCPCKRGVIGRCODRCNDNPFAYTNGCE 1971  
 RESULT 7  
 ID ABB11404 standard; peptide; 2560 AA.  
 AC ABB11404;  
 DT 11-JAN-2002 (first entry)  
 DE Human FLAMINGO 1 homologue, SEQ ID NO:11774.

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; actin; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; anti-inflammation;  
 KW antiasthmatic; antiarthritis; haemostatic; antiatherosclerotic;  
 KW cytoskeletal; osteopathic; vasodilator; cardiac; virucide; antibacterial;  
 KW antifungal; vulnery; antitumor.  
 OS Homo sapiens.  
 XX  
 XX W0200157188-A2.  
 PN 09-AUG-2001.  
 PD 05-FEB-2001, 2001WO-US03800.  
 PF 03-FEB-2000, 2000US-0496914.  
 PR 27-APR-2000, 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 PI WPI, 2001-457740/49.  
 DR N-PSDB; ABA08648.  
 DR  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 PS  
 PS Claim 20; Page 179-181, 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug

CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX Sequence 2560 AA;

Query Match 65.6%; Score 1917; DB 22; Length 2560;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2317; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	380	STTAATFLSVEDINDNAPQSEKRYVQVREDTGAPVLRATASDRDGSNAVHYXIM	439
DB	13	STTAATFLSVEDINDNAPQSEKRYVQVREDTGAPVLRATASDRDGSNAVHYXIM	72
QY	440	SGNARQFTLADAGTALDVVSPLDYETKXETTLRAADGGRPLSNVSGLVTVQVLDIN	499
DB	73	SGNARQFTLADAGTALDVVSPLDYETKXETTLRAADGGRPLSNVSGLVTVQVLDIN	132
QY	500	DNAIFVSTPQATVLESPVPLGLVLAHQALIDADAGNARLETRLGVGHDPFTINNGT	559
DB	133	DNAIFVSTPQATVLESPVPLGLVLAHQALIDADAGNARLETRLGVGHDPFTINNGT	192
QY	560	GNISVAELDRREVDYSGVEARDHGTPLTASASVSTVLDVNDNPFPTQPEYVRL	619
DB	193	GNISVAELDRREVDYSGVEARDHGTPLTASASVSTVLDVNDNPFPTQPEYVRL	252
QY	620	NEDAAVGVSTVSAVDRAHSAVITYQTSGNTRNPSITSGGGVLSIALPLDYKLER	679
DB	253	NEDAAVGVSTVSAVDRAHSAVITYQTSGNTRNPSITSGGGVLSIALPLDYKLER	312
QY	680	QVLAATASDGTRODPAQIVNANTDANTHRPVPQSHYTVNVEDRPAGTTVVLSATDE	739
DB	313	QVLAATASDGTRODPAQIVNANTDANTHRPVPQSHYTVNVEDRPAGTTVVLSATDE	372
QY	740	DTGENARITYFMEDSIPOFRIDADTGAVTQAELEDVDSYTLATTAARDNGIPQSDYT	799
DB	373	DTGENARITYFMEDSIPOFRIDADTGAVTQAELEDVDSYTLATTAARDNGIPQSDYT	432
QY	800	YIEILVNDVNDNAPOFLRDSYQSGVTEDEVPPPTSVLQISATDRDGLNGRVYTFQGGD	859
DB	433	YIEILVNDVNDNAPOFLRDSYQSGVTEDEVPPPTSVLQISATDRDGLNGRVYTFQGGD	492
QY	860	GGGDFIVESTSGIVRLRLRLRENVAAQVLAAYNDKMPARTPMEVTVVTLVDVNDP	919
DB	493	GGGDFIVESTSGIVRLRLRLRENVAAQVLAAYNDKMPARTPMEVTVVTLVDVNDP	552
QY	920	VEODEFDVFEVENSPIGLAVARVTAATDDEGTNAQIMQIIEGNIPEYFOLDIFSGELT	979
DB	553	VEODEFDVFEVENSPIGLAVARVTAATDDEGTNAQIMQIIEGNIPEYFOLDIFSGELT	612
QY	980	ALVLDLDYEDRPVYLVIAQTSAPLVSRATVHVELLDNRNDPVLGNFEILLNNYVTRSS	1039
DB	613	ALVLDLDYEDRPVYLVIAQTSAPLVSRATVHVELLDNRNDPVLGNFEILLNNYVTRSS	672
QY	1040	SPPGGAIGVPAHDDPISLSTYSPERGHELSIVLNASTGELKSRALDNNRPLEAINS	1099
DB	673	SPPGGAIGVPAHDDPISLSTYSPERGHELSIVLNASTGELKSRALDNNRPLEAINS	732
QY	1100	VLAUSDVHVSVAQALRVTITIDEMLTSHITLRLJEDMSPERFAPLGLFOAVATLAT	1159
DB	733	VLAUSDVHVSVAQALRVTITIDEMLTSHITLRLJEDMSPERFAPLGLFOAVATLAT	792
QY	1160	PPDHVVVENVQRTDAPGGHILNLSISVQPPGPGGPPPLPSEDLQERLYNRSLLTAI	1219
DB	793	PPDHVVVENVQRTDAPGGHILNLSISVQPPGPGGPPPLPSEDLQERLYNRSLLTAI	852
QY	1220	SAORVLPDPDNI CLREPCENYRCSVLA FPDSSAPRTASSVLPFPIHVVGGLRCRCPPG	1279
DB	853	SAORVLPDPDNI CLREPCENYRCSVLA FPDSSAPRTASSVLPFPIHVVGGLRCRCPPG	912
QY	1280	FTGDVCEFEVDLCYSRPGPHGRCSREGGYTCLCRDGYTGHCHEVSARSGRCTPGVCN	1339
DB	913	FTGDVCEFEVDLCYSRPGPHGRCSREGGYTCLCRDGYTGHCHEVSARSGRCTPGVCN	972

QY	1340	GGTCVNLLVGGFKCDPCPSGDFEKPYCOVTRSRPASHPTFRGLRQRFHTLSPATKE	1399
DB	973	GGTCVNLLVGGFKCDPCPSGDFEKPYCOVTRSRPASHPTFRGLRQRFHTLSPATKE	1032
QY	1400	RDGILLNKRPNKEDPVALEYIOBOVOLTEBAGSSTTVSPFPGVSDGOMHTVQLKY	1459
DB	1033	RDGILLNKRPNKEDPVALEYIOBOVOLTEBAGSSTTVSPFPGVSDGOMHTVQLKY	1092
QY	1460	YKRPILLGOTGLQSGSEOKVAVTVDCDITVLAALPGSVLAGYNSCAQGTQGGSKSLDL	1519
DB	1093	YKRPILLGOTGLQSGSEOKVAVTVDCDITVLAALPGSVLAGYNSCAQGTQGGSKSLDL	1152
QY	1520	TGPILLGVPDLPSFPVRMRQFVGCMTLQVDSRHIMADFIANNITVPGCPAKKNVCD	1579
DB	1153	TGPILLGVPDLPSFPVRMRQFVGCMTLQVDSRHIMADFIANNITVPGCPAKKNVCD	1212
QY	1580	SNYCHNGGTCVNQWMAFSCCEPLGFGKSCAQEMANPQHFGLSSLVAMHGLSLPISQPMY	1639
DB	1213	SNYCHNGGTCVNQWMAFSCCEPLGFGKSCAQEMANPQHFGLSSLVAMHGLSLPISQPMY	1272
QY	1640	LSIMFRTQADGVLLQAITRGRSTITLQIREGHVNLSTEGTGLQASSLRLBRGRANDGM	1699
DB	1273	LSIMFRTQADGVLLQAITRGRSTITLQIREGHVNLSTEGTGLQASSLRLBRGRANDGM	1332
QY	1700	HHAQALGASGGPGHAILSPDYGOQRAEAGLRLHGLHLSNITVGGIPGAGVAKGR	1759
DB	1333	HHAQALGASGGPGHAILSPDYGOQRAEAGLRLHGLHLSNITVGGIPGAGVAKGR	1392
QY	1760	GGLQGVRSIDTPREGNSLDPSHGESINVEGCSLPDPCDSNCPANSTCSNDWDSYSGC	1819
DB	1393	GGLQGVRSIDTPREGNSLDPSHGESINVEGCSLPDPCDSNCPANSTCSNDWDSYSGC	1452
QY	1820	DPGYGDNCNVCDLNPCEHOSVCTRKPSAPHYGCECPENYLGFCETRIDQPPCRGM	1879
DB	1453	DPGYGDNCNVCDLNPCEHOSVCTRKPSAPHYGCECPENYLGFCETRIDQPPCRGM	1512
QY	1880	GHPGCPNCCVSKGPDPCNKTSGBCHCKENHYRPPSPPTCLDCYPTGSLSRVCDPE	1939
DB	1513	GHPGCPNCCVSKGPDPCNKTSGBCHCKENHYRPPSPPTCLDCYPTGSLSRVCDPE	1572
QY	1940	DGQPCCKRGVIGROCDRCNPPABVTNMGCEVYNDSCEPRAIEAGTWMPRTRGLEPAAAPC	1999
DB	1573	DGQPCCKRGVIGROCDRCNPPABVTNMGCEVYNDSCEPRAIEAGTWMPRTRGLEPAAAPC	1632
QY	2000	PKGSFGTAVRHCDERGMPLPMLFNCTSIITSELGKFBRLQORNSGJDSGSOQALILL	2059
DB	1633	PKGSFGTAVRHCDERGMPLPMLFNCTSIITSELGKFBRLQORNSGJDSGSOQALILL	1692
QY	2060	RNATQHTAGYFGSDVKVAYOLATRLLAHESIQRGFGLSATODVHFTENILRVGSLDDTA	2119
DB	1693	RNATQHTAGYFGSDVKVAYOLATRLLAHESIQRGFGLSATODVHFTENILRVGSLDDTA	1752
QY	2120	NKRHWELIQOTBEGTAMLLQHTBEAYASLLAONMHTTISPTTIYTPNIVISVRLDKNF	2179
DB	1753	NKRHWELIQOTBEGTAMLLQHTBEAYASLLAONMHTTISPTTIYTPNIVISVRLDKNF	1812
QY	2180	AGAKLPRYBALRGSDPDLSTTVLLPESVFPETPPVVPAGBGEAOBEBELARRRDE	2239
DB	1813	AGAKLPRYBALRGSDPDLSTTVLLPESVFPETPPVVPAGBGEAOBEBELARRRDE	1872
QY	2240	LSQGEAVASVITTYTLAGLPHANTDPDRSLRVPKRP1INTPVVSI VYHDEBELPRALD	2299
DB	1873	LSQGEAVASVITTYTLAGLPHANTDPDRSLRVPKRP1INTPVVSI VYHDEBELPRALD	1932
QY	2300	KPVTVQFPLTEBETKPICVFNMHSIIVSGTGMASAGCEVFPNESHVSCQCHMTSF	2359
DB	1933	KPVTVQFPLTEBETKPICVFNMHSIIVSGTGMASAGCEVFPNESHVSCQCHMTSF	1992
QY	2360	AVLMDVSRRENGELPLKTLTYVALGVTLAALLTFPFTLLRLIRSNQHGIRNLTAL	2419
DB	1993	AVLMDVSRRENGELPLKTLTYVALGVTLAALLTFPFTLLRLIRSNQHGIRNLTAL	2052
QY	2420	GLAQVLELGINQADLPACTVIAILLHFLYLCFSPWALLRALHLYRALTEVRDVTGPM	2479

Db 2053 GLAQLVFLGINQADLPFACVLAIIILHPIYLCTFSSWALLLEALILYALTEVEDVTGPM 2112  
 Qy 2480 RFFYMLGMGPALFTGLAVGLDPRGVGNPDRCWMLSTYDTLIMSAGVAVAVNSVLTLY 2539  
 Db 2113 RFFYMLGMGPALFTGLAVGLDPRGVGNPDRCWMLSTYDTLIMSAGVAVAVNSVLTLY 2172  
 Qy 2540 LAARASCAAROGFEKKGPVSGLOPSFAVLLLSATWLLLSVNSDTLLFHYLFATCNC 2599  
 Db 2173 LAARASCAAROGFEKKGPVSGLOPSFAVLLLSATWLLLSVNSDTLLFHYLFATCNC 2232  
 Qy 2600 IQGFPIFLSYVLSKEYRKALKACSRKPSDDPALTTKSTLTSYNSCPSPYADRLYQPY 2659  
 Db 2233 IQGFPIFLSYVLSKEYRKALKACSRKPSDDPALTTKSTLTSYNSCPSPYADRLYQPY 2292  
 Qy 2660 GDSAGSLHSTRSGSKQPSYIPPLIRESSALNPQGGPGIG 2700  
 Db 2293 GDSAGSLHSTRSGSKQPSYIPPLIRESSALNPQGGPGIG 2333

## RESULT 8

AAB42192

ID AAB42192. standard; Protein: 2405 AA.

AC AAB42192;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF1956 polypeptide sequence SEQ ID NO:3912.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KM vulnery; antiposrotic; antiparkinsonian; noctropic; neuroprotective;  
 KM anticovulsant; osteoporotic; antichrilitic; immunosuppressant; candidant;  
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KM antiviral; antibacterial; antifungal; antineumatic; antichyroid;  
 KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KM cholesterol ester storage; systemic lupus erythematosus; infection;  
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KM thrombosis; contraceptive.

OS Homo sapiens.

XX MO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC76401.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 3067-3072; 5507P; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

XX which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiposrotic; antiparkinsonian; noctropic; neuroprotective;  
 CC osteoporotic; antichrilitic; immunosuppressant;  
 CC immunostimulant; candidant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antichyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 2405 AA;

Query Match 55.5%; Score 1623; DB 21; Length 2405;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1301 GRCSRREGYTCLCRGYTGHECEVARSGRCTPGYCKNGTGNLLVGGFKDCPSGDF 1360  
 Db 783 GRCSRREGYTCLCRGYTGHECEVARSGRCTPGYCKNGTGNLLVGGFKDCPSGDF 842  
 Qy 1361 EKPVCVTRSPFPAHSFIFRGLRQRFHTLASFAFKERDGLLYNGRFNEKHDFVALE 1420  
 Db 843 EKPVCVTRSPFPAHSFIFRGLRQRFHTLASFAFKERDGLLYNGRFNEKHDFVALE 902  
 Qy 1421 VIQBOVLTPSAGESTTYSPPVPGVSDGQWHTVQLKYNRPLLDQTLPGQSPSQKXA 1480  
 Db 903 VIQBOVLTPSAGESTTYSPPVPGVSDGQWHTVQLKYNRPLLDQTLPGQSPSQKXA 962  
 Qy 1481 VVTVDCDTPVALRFSVLYGNSCAAGTQGGSKSLDTGLPLAGVDDLPESFVVRMR 1540  
 Db 963 VVTVDCDTPVALRFSVLYGNSCAAGTQGGSKSLDTGLPLAGVDDLPESFVVRMR 1022  
 Qy 1541 QFVGCNRNLQVDSRHIDMADFLANNGTVPQCPAKXVCSNTCHNGTCVYNQMDAFSCC 1600  
 Db 1023 QFVGCNRNLQVDSRHIDMADFLANNGTVPQCPAKXVCSNTCHNGTCVYNQMDAFSCC 1082  
 Qy 1601 PLGFGKSCAQEMANPOHPLGSLVAMHGLSLPISQPYLISLMPFRROADGVLLQAITRG 1660  
 Db 1083 PLGFGKSCAQEMANPOHPLGSLVAMHGLSLPISQPYLISLMPFRROADGVLLQAITRG 1142  
 Qy 1661 RSTTTQLREGHVMSVEGTGLQASSLRLEPGRANDGWHHAQLAGSGPGHAILISFD 1720  
 Db 1143 RSTTTQLREGHVMSVEGTGLQASSLRLEPGRANDGWHHAQLAGSGPGHAILISFD 1202  
 Qy 1721 YGQARAGNLGPRLLHGLHLSNITVGGIPLPAGVARGFGGCTQVAVSDTPBSGNSLDES 1780  
 Db 1203 YGQARAGNLGPRLLHGLHLSNITVGGIPLPAGVARGFGGCTQVAVSDTPBSGNSLDES 1262  
 Qy 1781 HGRSINVBQCSLPDPCDNPCCPANSYCSNDMYSQCDPGYGNCTNVCDLNPCEHQ 1840  
 Db 1263 HGRSINVBQCSLPDPCDNPCCPANSYCSNDMYSQCDPGYGNCTNVCDLNPCEHQ 1322  
 Qy 1841 SVCTRKPSAPHGTGCECPNVLGAPYCESTRIDQCPRGWGHPTGBCNCDVSKGPPDCN 1900  
 Db 1323 SVCTRKPSAPHGTGCECPNVLGAPYCESTRIDQCPRGWGHPTGBCNCDVSKGPPDCN 1382  
 Qy 1901 KTSGECHKENHRRPPGSPFCLCDCTPGSLSRVCDPBGQCPCKPVYGRQCDKCDNP 1960  
 Db 1383 KTSGECHKENHRRPPGSPFCLCDCTPGSLSRVCDPBGQCPCKPVYGRQCDKCDNP 1442  
 Qy 1961 FAEVTNGCEVNVDSGCPRAIEAGIWWPRTFRGLPAAPCKGSGFGTAVRHCDHRGMLP 2020  
 Db 1443 FAEVTNGCEVNVDSGCPRAIEAGIWWPRTFRGLPAAPCKGSGFGTAVRHCDHRGMLP 1502



```

OY 2021 NLFNCTSTTFSEELKCPAERLQNEGSLDGRSCQQLALLRNATQHTAGYFGSDVAVAYQL 2080
DB 1503 NLFNCTSTTFSEELKCPAERLQNEGSLDGRSCQQLALLRNATQHTAGYFGSDVAVAYQL 1562
OY 2081 ATRLIAHSTORGFSLATQDVHFTENILLRVSALLDTANRKMELIQTEGTWMLLOH 2140
DB 1563 ATRLIAHSTORGFSLATQDVHFTENILLRVSALLDTANRKMELIQTEGTWMLLOH 1622
OY 2141 YEAVASALAQNMRHTYLSPTITVTNIVISVRLDKGNFAGAKLPRYEALRGEOPDLET 2200
DB 1663 YEAVASALAQNMRHTYLSPTITVTNIVISVRLDKGNFAGAKLPRYEALRGEOPDLET 1682
OY 2201 TVILPESVFRETTPVVRPAPGEAQEPBEELARQRHPELSQGBAVASYIYRTLAGLLP 2260
DB 1683 TVILPESVFRETTPVVRPAPGEAQEPBEELARQRHPELSQGBAVASYIYRTLAGLLP 1742
OY 2261 HNYDDKXSLRVPKRPINTPVPVSTSYHDEELLPRALDKPTVQFRLLETERKPCV 2320
DB 1743 HNYDDKXSLRVPKRPINTPVPVSTSYHDEELLPRALDKPTVQFRLLETERKPCV 1802
OY 2321 FMNHSILVSGTGMGARGCEVFRNESHVSQCNMTSPAVLMDVSRRENGEILPLKTLT 2380
DB 1803 FMNHSILVSGTGMGARGCEVFRNESHVSQCNMTSPAVLMDVSRRENGEILPLKTLT 1862
OY 2381 YVALGVTLAALLTFEFLTLRLRSNOGIRRNITAAAGLAQVFLGINDLPPACT 2440
DB 1863 YVALGVTLAALLTFEFLTLRLRSNOGIRRNITAAAGLAQVFLGINDLPPACT 1922
OY 2441 VIATILHFLYCTFSWALLBALHLYRALTEVRDVNTGPMRFTYMLGNGVPAITGLANGL 2500
DB 1923 VIATILHFLYCTFSWALLBALHLYRALTEVRDVNTGPMRFTYMLGNGVPAITGLANGL 1982
OY 2501 DPEGYNDFPCMLSIYDTLINSFAPVAVASMSVFLYTLAARASCAAROGFEKGPVS 2560
DB 1983 DPEGYNDFPCMLSIYDTLINSFAPVAVASMSVFLYTLAARASCAAROGFEKGPVS 2042
OY 2561 GLOPSFAVLLLSATWLLALISVNSDTLLFHYLFATCNCIOGPFIFLSYVVLSEKVRAL 2620
DB 2043 GLOPSFAVLLLSATWLLALISVNSDTLLFHYLFATCNCIOGPFIFLSYVVLSEKVRAL 2102
OY 2621 KLAGCRKSPDPALTTKSTLTSVNCSPYADGRLYQYQYGSAGSLHSTSRGKQPEYI 2680
DB 2103 KLAGCRKSPDPALTTKSTLTSVNCSPYADGRLYQYQYGSAGSLHSTSRGKQPEYI 2162
OY 2681 PFLAREBSALNPGQPPGLGDPGSLFLEGGDQDHPDTSDSLSDDDQSGSYASTSS 2740
DB 2163 PFLAREBSALNPGQPPGLGDPGSLFLEGGDQDHPDTSDSLSDDDQSGSYASTSS 2222
OY 2741 DSEEBEEREBEBAAPGEGQWDSLIGPAGERLPLHSTPKDGGPGGKAPWPGDFGTAK 2800
DB 2223 DSEEBEEREBEBAAPGEGQWDSLIGPAGERLPLHSTPKDGGPGGKAPWPGDFGTAK 2282
OY 2801 SSGNGAPEREERENDALSRGSLGPLPGSSAOPHKGLIKKCLPTISEKSLRLPEQ 2860
DB 2283 SSGNGAPEREERENDALSRGSLGPLPGSSAOPHKGLIKKCLPTISEKSLRLPEQ 2342
OY 2861 CTGSSRGSASGSGSGPPPPPPRQSLQEBOLNGWPIAMSIIKAGTVDESSGSEFLPFN 2920
DB 2343 CTGSSRGSASGSGSGPPPPPPRQSLQEBOLNGWPIAMSIIKAGTVDESSGSEFLPFN 2402
OY 2921 FLH 2923
DB 2403 FLH 2405

```

```

XX DE Human calcitonin receptor.
XX XX
KM K Calictonin receptor; HCRPR64; G-protein coupled receptor; human;
KM K infection; migraine; cancer; anorexia; bulimia; asthma; allergy;
KM K Parkinson's disease; acute heart failure; hypertension;
KM K hypertension; urinary retention; osteoporosis; angina pectoris;
KM K myocardial infarction; ulcer; benign prostatic hypertrophy;
KM K psychosis; neurological diseases; anxiety; schizophrenia;
KM K manic depression; delirium; dementia; mental retardation;
KM K dyskinesia; Huntington's disease; Gilles de la Tourette syndrome;
KM K diagnosis; therapy.
XX XX
OS Homo sapiens.
XX XX
FH Key
FH Peptide
FT /label= S1g_peptide
FT /label= 1..16
FT /label= 1..18
FT /label= Extracellular
FT /note= "Claim 10"
FT /label= 20..49
FT /label= Transmembrane_domain-1
FT /note= "Claim 10"
FT /label= 50..59
FT /label= Intracellular
FT /note= "Claim 10"
FT /label= 60..80
FT /label= Transmembrane_domain-2
FT /note= "Claim 10"
FT /label= 82..109
FT /label= Transmembrane_domain-3
FT /note= "Claim 10"
FT /label= 110..125
FT /label= Intracellular
FT /note= "Claim 10"
FT /label= 126..147
FT /label= Transmembrane_domain-4
FT /note= "Claim 10"
FT /label= 148..159
FT /label= Extracellular
FT /note= "Claim 10"
FT /label= 160..187
FT /label= Transmembrane_domain-5
FT /note= "Claim 10"
FT /label= 188..209
FT /label= Intracellular
FT /note= "Claim 10"
FT /label= 209..230
FT /label= Transmembrane_domain-6
FT /note= "Claim 10"
FT /label= 231..236
FT /label= Extracellular
FT /note= "Claim 10"
FT /label= 237..258
FT /label= Transmembrane_domain-7
FT /note= "Claim 10"
FT /label= 259..560
FT /label= Intracellular
FT /note= "Claim 10"
FT /label= 1..18
FT /label= Extracellular
FT /note= "Claim 10"
XX PN WO9821242-A1.
XX XX
XX PD 22-MAY-1998.
XX PF 14-NOV-1997;
XX PR 15-NOV-1996;
XX PA 97WO-US21330.
XX PA 96US-0030934.
XX PA (HUMA-) HUMAN GENOME SCI INC.

```

XX LI Y, Ruben SM, Soppet DR,  
 XX MPI; 1998-297869/26.  
 DR N-PSDB; AAV07219.  
 XX  
 PT New nucleic acid encoding calcitonin receptor or its fragments -  
 PT useful for diagnosis and treatment of infections, cancer, allergy or  
 PT neurological disease  
 PS  
 PS Claim 1; Fig 1A-B; 88pp; English.  
 CC  
 CC This polypeptide comprises a novel human calcitonin receptor (CR),  
 CC which is a member of the G-protein coupled receptor superfamily.  
 CC Its amino acid sequence was deduced from a human cerebellum cDNA  
 CC clone (see AAV07219). The receptor is about 21% identical to rat  
 CC CR and about 31% identical to hormone receptor EMRI. CR  
 CC polypeptides, including extracellular, intracellular and  
 CC transmembrane regions, as well as epitope-bearing portions, are  
 CC claimed, as are isolated nucleic acids, recombinant vectors, host  
 CC cells and antibodies. Conditions related to under-expression or  
 CC over-expression of calcitonin receptor are treated by administering  
 CC an (ant)agonist of the receptor. Specified conditions are infections  
 CC (bacterial, fungal, protozoal or viral, particularly HIV-1 or  
 CC HIV-2), pain, e.g. migraine, cancer, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypertension, hypotension,  
 CC urinary retention, osteoporosis, angina pectoris, myocardial  
 CC infarction, ulcer, allergy, benign prostatic hypertrophy, psychotic  
 CC and neurological diseases (anxiety, schizophrenia, manic depression,  
 CC delirium, dementia and severe mental retardation) and dyskinesia  
 CC (e.g. Huntington's disease or Gilles de la Tourette syndrome) (all  
 CC claimed). Host cells are used to express recombinant calcitonin  
 CC receptor polypeptides which are used as mol.wt. markers, as  
 CC immunogens, for isolation of interacting proteins (or their genes)  
 CC and to screen for therapeutic (ant)agonists.  
 XX  
 XX Sequence 568 AA;  
 SQ  
 Query Match 18.3%; Score 534; DB 19; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2388 LAALLTFPFLTLIRLSNQHGRRLTAALGAQVYLLGNODLPACVIALILH 2447  
 DB 33 LAALLTFPFLTLIRLSNQHGRRLTAALGAQVYLLGNODLPACVIALILH 92  
 QY 2448 FLYLCTFSMALLEALHLRYALTVEADVNTGPMRFYVLMGCVAFITGLAVGDPGSGN 2507  
 DB 93 FLYLCTFSMALLEALHLRYALTVEADVNTGPMRFYVLMGCVAFITGLAVGDPGSGN 152  
 QY 2508 PDECFWLSIYDTLIMSFPAGVAFVAVSMVFYLLAARASCAAROGPEKGPVSGLOPSFA 2567  
 DB 153 PDECFWLSIYDTLIMSFPAGVAFVAVSMVFYLLAARASCAAROGPEKGPVSGLOPSFA 212  
 QY 2568 VLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSYVVLSEKVKAKLACSRK 2627  
 DB 213 VLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSYVVLSEKVKAKLACSRK 272  
 QY 2628 PSEDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGLSHSTRSGKQSPYIFLLRRE 2687  
 DB 273 PSEDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGLSHSTRSGKQSPYIFLLRRE 332  
 QY 2688 SAANPGCGPGLDPSGLFLBGOQOQHDPTDSDSLSEDDSGSYASTHSSDSEERE 2747  
 DB 333 SAANPGCGPGLDPSGLFLBGOQOQHDPTDSDSLSEDDSGSYASTHSSDSEERE 392  
 QY 2748 EEEBEAFAFPGEGWDSLLGFGAERLPLHSTPKDGGPGPKAPMGDGGTAKSSGNGAP 2807  
 DB 393 EEEBEAFAFPGEGWDSLLGFGAERLPLHSTPKDGGPGPKAPMGDGGTAKSSGNGAP 452  
 QY 2808 EERLRNGDALSRGSLGAPLPGSSAQPHKGLKKKCLPTISKSSLLRLPLECOTGSSRG 2867  
 DB 453 EERLRNGDALSRGSLGAPLPGSSAQPHKGLKKKCLPTISKSSLLRLPLECOTGSSRG 512

QY 2868 SSASBSRGPPPPRPPROSLQEOINGVPIAMSIKAGTVDESSGSEFLPFNF 2921  
 DB 513 SSASBSRGPPPPRPPROSLQEOINGVPIAMSIKAGTVDESSGSEFLPFNF 566  
 RESULT 10  
 ABJ37074  
 ID ABJ37074 standard; Protein; 565 AA.  
 XX  
 XX ABJ37074;  
 AC  
 XX 01-MAY-2003 (first entry)  
 DT  
 XX Human breast cancer / ovarian cancer related protein #50.  
 DE  
 XX Human; cytostatic; breast cancer; ovarian cancer.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO2003000012-A2.  
 PN  
 XX 03-JAN-2003.  
 PD  
 XX 21-JUN-2002; 2002MO-US19773.  
 PF  
 XX 21-JUN-2001; 2001US-300159P.  
 PR 27-JUN-2001; 2001US-301351P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX Velby OP;  
 PI  
 XX MPI; 2003-267848/26.  
 DR N-PSDB; ABT31943.  
 XX  
 PT Determining the presence of breast cancer in an individual, involves  
 PT using specific polynucleotide markers -  
 PS Disclosure; Page 219-220; 233pp; English.  
 XX  
 CC The invention comprises a method for assessing whether a patient is  
 CC afflicted with breast cancer or ovarian cancer. The method involves the  
 CC use of specific DNA markers. The method of the invention is useful in the  
 CC detection and treatment of ovarian and breast cancer. Amino acid  
 CC sequences ABJ37025 - ABJ37080 represent human breast/ovarian cancer-  
 CC related proteins.  
 CC  
 XX  
 XX Sequence 565 AA;  
 SQ  
 Query Match 16.6%; Score 485; DB 24; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2430 INQADLPFCYVIALILHLYICTFSMALLEALHLRYALTVEADVNTGPMRFYVLMGCV 2489  
 DB 74 INQADLPFCYVIALILHLYICTFSMALLEALHLRYALTVEADVNTGPMRFYVLMGCV 133  
 QY 2490 PAFITGLAVGLDPBGXGNDPFCMLSIYDTLIMSFPAGVAFVAVSMVFYLLAARASCAQ 2549  
 DB 134 PAFITGLAVGLDPBGXGNDPFCMLSIYDTLIMSFPAGVAFVAVSMVFYLLAARASCAQ 193  
 QY 2550 KQGFEEKGVSGLQPSFAVLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSY 2609  
 DB 194 KQGFEEKGVSGLQPSFAVLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSY 253  
 QY 2610 VVLSEKVKAKLACSRKSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGLSHST 2669  
 DB 254 VVLSEKVKAKLACSRKSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGLSHST 313  
 QY 2670 SRSKQSPSYIFLLRREBALNPGCGPGLDPSGLFLBGOQOQHDPTDSDSLSEDD 2729  
 DB 314 SRSKQSPSYIFLLRREBALNPGCGPGLDPSGLFLBGOQOQHDPTDSDSLSEDD 373



CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytostatic,  
 CC cardioactive, immunomodulatory, muscular, vulnerability, gastrointestinal,  
 CC nephrotoxic, anti-infective, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF15506 to AAF15514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.  
 CC  
 XX Sequence 717 AA;  
 SQ  
 Query Match 9.6%; Score 281; DB 21; Length 717;  
 Best Local Similarity 100.0%; Pred. No. 2,2e-249; Indels 0; Gaps 0;  
 Matches 281; Conservative 0; Mismatches 0;  
 QY 2596 TCNCIGGPFILSYVLSKVRKALKACSRKSPDPALTTKSTLTSSVNCPPYADGRL 2655  
 DB 390 TCNCIGGPFILSYVLSKVRKALKACSRKSPDPALTTKSTLTSSVNCPPYADGRL 449  
 QY 2656 YQPYGDSAGSLHSTSRSGKQPSYIPLREBSALNPGQGPGLGDPGSLFLGQDQGH 2715  
 DB 450 YQPYGDSAGSLHSTSRSGKQPSYIPLREBSALNPGQGPGLGDPGSLFLGQDQGH 509  
 QY 2716 PPDSDSLSLDDQSGSYASTSSDSEEESEEESEEESEEAAPRGCGMDSLGPAERLPLH 2775  
 DB 510 PPDSDSLSLDDQSGSYASTSSDSEEESEEESEEESEEAAPRGCGMDSLGPAERLPLH 569  
 QY 2776 STPKDGGPGGKAPWPGDPFTTAKSSGNGAPERLRENDALSRGSLGPLPGSSAQPH 2835  
 DB 570 STPKDGGPGGKAPWPGDPFTTAKSSGNGAPERLRENDALSRGSLGPLPGSSAQPH 629  
 QY 2836 KGILKKKCLPTISEKSLRLPLBQCTGSSRGSSASGSRG 2876  
 DB 630 KGILKKKCLPTISEKSLRLPLBQCTGSSRGSSASGSRG 670  
 QY  
 DB  
 RESULT 13  
 ABBP5877  
 ID ABBP5877 standard; Protein; 470 AA.  
 XX  
 AC ABBP5877;  
 XX  
 DT 10-FEB-2003 (first entry)  
 XX  
 DE Human secretory polypeptide SPTM SEQ ID NO 1061.  
 XX  
 KW Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;  
 KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;  
 KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;  
 KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;  
 KW anti-inflammatory; immunosuppressive; neuroprotective; nootropic;  
 KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; antidiabetic;  
 KW antineoplastic; antianemic; anti-HIV; human immunodeficiency virus;  
 KW secretory polynucleotide; secretory protein.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283876-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09921.  
 XX  
 PR 29-MAR-2001; 2001US-28067P.  
 PR 29-MAR-2001; 2001US-28068P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 17-MAY-2001; 2001US-291849P.

PR 19-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-30001P.  
 XX  
 PA (INCY- ) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J,  
 PI DuFour GE, Hillman JL, Yu JY, Tuason O, Yap PB, Amshy SR,  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH,  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashba ME;  
 XX  
 DR WPI; 2003-075543/07.  
 DR N-PSDB; AB236319.  
 XX  
 PT New human secretory proteins and polynucleotides, useful for  
 PT diagnosing, treating or preventing autoimmune/inflammatory disorders  
 PT (e.g. AIDS), neurological disorders (e.g. Alzheimer's), or cell  
 PT proliferations or cancers -  
 XX  
 PS Claim 27; SEQ ID NO 1061; 458pp + Sequence Listing; English.  
 XX  
 CC The invention relates to a secretory polynucleotide (designated sptm)  
 CC comprising any of 567 polynucleotide sequences (AB235837-AB236402), a  
 CC naturally occurring polynucleotide sequence at least 90 % identical to  
 CC the polynucleotide sequence, a polynucleotide complementary to them or an  
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for  
 CC treating, preventing or diagnosing a disease or condition associated with  
 CC the expression of functional SPTM. These are particularly useful for  
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders  
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's  
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,  
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,  
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,  
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.  
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,  
 CC breast, cervix or prostate). The present sequence is one of the SPTM  
 CC proteins of the invention (ABP75384-ABP75962).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 470 AA;  
 XX  
 Query Match 7.5%; Score 220; DB 24; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-193; Indels 0; Gaps 0;  
 Matches 220; Conservative 0; Mismatches 0;  
 QY 2560 SGLQSPFAVLLLSATWLLALLSVNSDTLLFHYLFATNCIGPPIFLSYVLSKVRKA 2619  
 DB 36 SGLQSPFAVLLLSATWLLALLSVNSDTLLFHYLFATNCIGPPIFLSYVLSKVRKA 95  
 QY 2620 LKLACSRKSPDPALTTKSTLTSSVNCPPYADGSLYQPYGDSAGSLHSTSRSGKQPSY 2679  
 DB 96 LKLACSRKSPDPALTTKSTLTSSVNCPPYADGSLYQPYGDSAGSLHSTSRSGKQPSY 155  
 QY 2680 IFFLLREBSALNPGQGPGLGDPGSLFLGQDQGHDPDSDSLSLDDQSGSYASTHS 2739  
 DB 156 IFFLLREBSALNPGQGPGLGDPGSLFLGQDQGHDPDSDSLSLDDQSGSYASTHS 215  
 QY 2740 SDSEEESEEESEEAAPRGCGMDSLGPAERLPLHSTPK 2779  
 DB 216 SDSEEESEEESEEAAPRGCGMDSLGPAERLPLHSTPK 255  
 QY  
 DB  
 RESULT 14  
 ABB03070  
 ID ABB03070 standard; Protein; 219 AA.  
 XX  
 AC ABB03070;  
 XX  
 DT 07-JAN-2002 (first entry)

XX DE Human expressed polypeptide SEQ ID NO 43.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;  
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; human; secreted protein.  
XX OS Homo sapiens.  
XX PN WO200155167-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01319.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 26-SEP-2000; 2000US-0235484.  
XX PR 01-DEC-2000; 2000US-0250160.  
XX PR 05-DEC-2000; 2000US-0251988.  
XX PR 06-DEC-2000; 2000US-0251479.  
XX PR 08-DEC-2000; 2000US-0251990.  
XX PR 11-DEC-2000; 2000US-0254097.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465559/50.  
XX N-PSDB; AA199622.  
XX PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
XX PT diagnose diseases or disorders associated with polypeptides and for  
XX PT treating autoimmune diseases e.g., multiple sclerosis, rheumatoid  
XX PT arthritis -  
XX PS Claim 11; SEQ ID NO 43; 504bp + Sequence Listing; English.  
XX CC The invention relates to novel genes (AA199614-AA199654) and proteins  
XX CC (AB03062-AB03065) useful for preventing, treating or ameliorating  
XX CC medical conditions e.g. by protein or gene therapy. The genes are  
XX CC isolated from a range of human tissues disclosed in the specification.  
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX CC and parasitic infections.  
XX CC Note: The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SO Sequence 219 AA;  
Query Match 4.0%; Score 118; DB 22; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.2e-99;  
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
567 ELDREVDYFSGVARDGCTALTRASASVTVLDVNNNNPTFQPEYTVLNDPAVG 626  
Db 1 ELDREVDYFSGVARDGCTALTRASASVTVLDVNNNNPTFQPEYTVLNDPAVG 60  
QY 627 TSVVTVSAVDRAHSAVITQTISGNTFRNFSTISGSGGLVSLALPLDKLRRQVYLA 684  
DB 61 TSVVTVSAVDRAHSAVITQTISGNTFRNFSTISGSGGLVSLALPLDKLRRQVYLA 118

RESULT 15  
AAU19681  
ID AAU19681 standard; Protein; 219 AA.  
XX AC AAU19681;  
XX DT 06-DEC-2001 (first entry)  
XX DE Human novel extracellular matrix protein, Seq ID No 331.  
XX KW Human; secreted extracellular matrix protein; immunomodulatory;  
XX KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;  
XX KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytosstatic;  
XX KW Alzheimer's disease; immune/autoimmune disease; HIV infection; anaemia;  
XX KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
XX KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;  
XX KW Senary syndrome; Gaucher's disease; neurological diseases;  
XX KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
XX KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
XX KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX OS Homo sapiens.  
XX PN WO200155368-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01348.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 11-JUL-2000; 2000US-0216880.  
XX PR 11-JUL-2000; 2000US-0217487.  
XX PR 14-JUL-2000; 2000US-0217496.  
XX PR 14-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 26-JUL-2000; 2000US-0220964.  
XX PR 14-AUG-2000; 2000US-0224518.  
XX PR 14-AUG-2000; 2000US-0224519.  
XX PR 14-AUG-2000; 2000US-0225213.  
XX PR 14-AUG-2000; 2000US-0225214.  
XX PR 14-AUG-2000; 2000US-0225266.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225268.  
XX PR 14-AUG-2000; 2000US-0225270.  
XX PR 14-AUG-2000; 2000US-0225447.  
XX PR 14-AUG-2000; 2000US-0225757.  
XX PR 14-AUG-2000; 2000US-0225758.  
XX PR 14-AUG-2000; 2000US-0225759.  
XX PR 18-AUG-2000; 2000US-0226279.  
XX PR 22-AUG-2000; 2000US-0226681.  
XX PR 22-AUG-2000; 2000US-0226868.  
XX PR 23-AUG-2000; 2000US-0227187.  
XX PR 23-AUG-2000; 2000US-0227009.  
XX PR 30-AUG-2000; 2000US-0228924.  
XX PR 01-SEP-2000; 2000US-0229287.  
XX PR 01-SEP-2000; 2000US-0229343.  
XX PR 01-SEP-2000; 2000US-0229344.  
XX PR 01-SEP-2000; 2000US-0229345.  
XX PR 05-SEP-2000; 2000US-0229509.  
XX PR 05-SEP-2000; 2000US-0229513.  
XX PR 06-SEP-2000; 2000US-0230437.

```

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234374.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

```

```

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250393.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

```

## (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM,  
WPI; 2001-465572/50.  
N-PSDB; AAG31252.

Nucleic acid molecules encoding human secreted extracellular matrix proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 11; SEQ ID No 331; 577pp; English.

The invention relates to isolated nucleic acid molecules encoding novel human secreted extracellular matrix proteins (SPs). The polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotides may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in samples. The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). Other uses include wound healing, maintenance of organs before transplantation.

Query Match 4.0%; Score 118; DB 22; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.2e-99; Indels 0; Gaps 0;

Matches 118; Conservative 0; Mismatches 0;

```

QY 567 ELDREKVDYFSGVGRARDHGTPTALTSASVSVTYADVNDNNPTFPQPEYTVRLNEDAAVG 626
DB 1 ELDREKVDYFSGVGRARDHGTPTALTSASVSVTYADVNDNNPTFPQPEYTVRLNEDAAVG 60
QY 627 TSVTVTSADVDRASHVITTYQITSGNTRNRFSTSGGGGLYSALPLPYKLERQYVLA 684
DB 61 TSVTVTSADVDRASHVITTYQITSGNTRNRFSTSGGGGLYSALPLPYKLERQYVLA 118

```



---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 11, 2004, 16:08:16 ; Search time 67 Seconds  
(without alignments)  
9134.693 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 2923  
Sequence: 1 MMSPTGVLPPTPPPLLL.....AGTVDEDSGSEPLFPNPLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 801455 seqs, 209382283 residues

Word size : 0

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/pubppa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubppa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubppa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/1/pubppa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubppa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubppa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubppa/US10C\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	2923	11	US-09-788-711A-4
2	2923	100.0	2923	11	US-09-916-849A-3
3	2923	100.0	2923	12	US-10-174-677-29
4	2923	100.0	2923	12	US-10-120-801-53
5	2923	100.0	2923	12	US-10-798-992-932
6	2923	100.0	2923	15	US-10-225-567A-524
7	2837	97.1	2956	9	US-09-788-711A-2
8	534	18.3	568	9	US-09-843-856-2
9	485	16.6	565	15	US-10-176-847-100
10	281	9.6	568	12	US-10-264-237-2041
11	281	9.6	717	10	US-09-925-300-1299
12	179	6.1	646	12	US-10-017-161-1096
13	118	4.0	219	9	US-09-764-870-331
14	118	4.0	219	15	US-10-125-540-331
15	100	3.4	111	9	US-09-764-870-479

16	100	3.4	111	15	US-10-125-540-479	Sequence 479, App
17	82	2.8	96	9	US-09-764-853-659	Sequence 659, App
18	82	2.8	96	9	US-09-764-896-286	Sequence 286, App
19	82	2.8	96	11	US-09-764-881-94	Sequence 94, App1
20	82	2.8	96	12	US-10-242-747-94	Sequence 94, App1
21	82	2.8	96	15	US-10-073-865-108	Sequence 108, App
22	59	2.0	174	12	US-10-017-161-1688	Sequence 1688, App
23	42	1.4	141	9	US-09-764-870-335	Sequence 335, App
24	42	1.4	141	15	US-10-125-540-335	Sequence 335, App
25	23	0.8	3034	9	US-09-737-149-25	Sequence 25, App1
26	23	0.8	3034	9	US-09-737-149-30	Sequence 30, App1
27	23	0.8	3034	12	US-10-131-409-70	Sequence 70, App1
28	23	0.8	3034	12	US-10-120-801-52	Sequence 52, App1
29	23	0.8	3034	12	US-10-150-811-70	Sequence 70, App1
30	21	0.7	44	12	US-09-880-573-16	Sequence 16, App1
31	20	0.7	20	11	US-09-916-849A-11	Sequence 11, App1
32	20	0.7	20	15	US-10-225-567A-186	Sequence 186, App
33	20	0.7	20	15	US-10-225-567A-1867	Sequence 1867, App
34	20	0.7	1713	9	US-09-737-149-27	Sequence 27, App1
35	20	0.7	3014	9	US-09-737-149-2	Sequence 2, App1
36	20	0.7	3014	12	US-10-241-220-107	Sequence 107, App
37	20	0.7	3014	12	US-10-174-677-77	Sequence 77, App1
38	20	0.7	3014	12	US-10-235-027-750	Sequence 750, App
39	20	0.7	3014	12	US-10-240-145-63	Sequence 63, App1
40	20	0.7	3014	15	US-10-225-567A-444	Sequence 444, App
41	20	0.7	3028	12	US-10-131-409-14	Sequence 14, App1
42	20	0.7	3028	12	US-10-131-409-69	Sequence 69, App1
43	20	0.7	3028	12	US-10-150-811-14	Sequence 14, App1
44	20	0.7	3028	12	US-10-150-811-69	Sequence 69, App1
45	19	0.7	19	15	US-10-225-567A-1865	Sequence 1865, App

## ALIGNMENTS

RESULT 1  
US-09-788-711A-4  
Sequence 4, Application US/09788711A  
Patent No. US20020058328A1  
GENERAL INFORMATION:  
APPLICANT: Tania Tamsin Testa  
TITLE OF INVENTION: NOVEL COMPOUNDS  
FILE REFERENCE: GR-30225  
CURRENT APPLICATION NUMBER: US/09/788, 711A  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 0004196.2  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 2923  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-788-711A-4

Query Match 100.0%; Score 2923; DB 9; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MMSPTGVLPPTPPPLLLLLPLLDGQVPCRSLSGRGSSGACAPMGLCP	60
DB	1	MMSPTGVLPPTPPPLLLLLPLLDGQVPCRSLSGRGSSGACAPMGLCP	60
QY	61	SASNTMLTSCRDRDGTETLVTGHLVPHDGLRWVCSESAHIFLPAPBGC	120
DB	61	SASNTMLTSCRDRDGTETLVTGHLVPHDGLRWVCSESAHIFLPAPBGC	120
QY	121	GHLSFGKLTLPPEHPCIKAPLRCSCTLAQAPFIRAGERSPESLSGRR	180
DB	121	GHLSFGKLTLPPEHPCIKAPLRCSCTLAQAPFIRAGERSPESLSGRR	180
QY	181	PPPSYQATVPENQAGTVPASIRAIIDDEGAGLETMTALPFSRSNQPSL	240
DB	181	PPPSYQATVPENQAGTVPASIRAIIDDEGAGLETMTALPFSRSNQPSL	240

Db 181 POPSYATVPENOPAGTVPASLRALIDPEDEGAEGLLEYTMDLPDRSNOFSLDPVTGA 240  
 Qy 241 VTTAEELDRKSTHFRVTAQDHGMPRRSALATITLITVTDNDHPVPEQOEKESLRE 300  
 Db 241 VTTAEELDRKSTHFRVTAQDHGMPRRSALATITLITVTDNDHPVPEQOEKESLRE 300  
 Qy 301 NLEVGVEVLVTRATDGAAPPNANILYRLLESGSGSPSEVFEBIDPRSGVIRTRGFVDEEV 360  
 Db 301 NLEVGVEVLVTRATDGAAPPNANILYRLLESGSGSPSEVFEBIDPRSGVIRTRGFVDEEV 360  
 Qy 361 ESYVOLTVEASDQDGPGRSTTAAVFLSVEDNDNNAPOFSKRVVVOVREVTGAYLR 420  
 Db 361 ESYVOLTVEASDQDGPGRSTTAAVFLSVEDNDNNAPOFSKRVVVOVREVTGAYLR 420  
 Qy 421 VTTASDRKGSNAVHYHISMSGNAGOFYLDQAGALDVSPFLDYETTKETYLTRAADGG 480  
 Db 421 VTTASDRKGSNAVHYHISMSGNAGOFYLDQAGALDVSPFLDYETTKETYLTRAADGG 480  
 Qy 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFGATVLESVPLGYLVAVQAIDADAGDNARL 540  
 Db 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFGATVLESVPLGYLVAVQAIDADAGDNARL 540  
 Qy 541 EYRLAGVGHDPFTINNGTGMISVAELDRREVDYFSGVEARDHGTPALTAASVSTV 600  
 Db 541 EYRLAGVGHDPFTINNGTGMISVAELDRREVDYFSGVEARDHGTPALTAASVSTV 600  
 Qy 601 LDVANDNPTFTQPEYTVRLNEDAVGTSVTVSAVDDAHSVITYQITSGNTRRFSITS 660  
 Db 601 LDVANDNPTFTQPEYTVRLNEDAVGTSVTVSAVDDAHSVITYQITSGNTRRFSITS 660  
 Qy 661 QSGGGLVSLALPLDYKLEROYVLAVTASDGTRODTAQIVNVTDANTHRPVGSHYTVN 720  
 Db 661 QSGGGLVSLALPLDYKLEROYVLAVTASDGTRODTAQIVNVTDANTHRPVGSHYTVN 720  
 Qy 721 VNEDRPACTTVVLSATDEDEGENARITYEMEDSIPOFRIDADGATTOAELDYEOVS 780  
 Db 721 VNEDRPACTTVVLSATDEDEGENARITYEMEDSIPOFRIDADGATTOAELDYEOVS 780  
 Qy 781 YTLAITARNDNGIPKSDPTTYLEILVNDVNDNAPOLFDRSDYQGSYVEDVPFTSVLQISAT 840  
 Db 781 YTLAITARNDNGIPKSDPTTYLEILVNDVNDNAPOLFDRSDYQGSYVEDVPFTSVLQISAT 840  
 Qy 841 DROSGLNGRYFTYQGGDGDGDFIVESTSGIVTTLRLDRBNVAQVYLRVAVDKMPP 900  
 Db 841 DROSGLNGRYFTYQGGDGDGDFIVESTSGIVTTLRLDRBNVAQVYLRVAVDKMPP 900  
 Qy 901 ARPMEVTVVTLVNDNPPVPEODEPVPVBEENSPIGLAVARVTAHPDDEGTNAQIMYOI 960  
 Db 901 ARPMEVTVVTLVNDNPPVPEODEPVPVBEENSPIGLAVARVTAHPDDEGTNAQIMYOI 960  
 Qy 961 VEGNIPPEVFOLDIFSGBELTALVLDVYEDREPEYVLVIQATSAPLVSRAVTHVRLDRNDP 1020  
 Db 961 VEGNIPPEVFOLDIFSGBELTALVLDVYEDREPEYVLVIQATSAPLVSRAVTHVRLDRNDP 1020  
 Qy 1021 PVUGNPEFILLNNYTTNRSSEFPGGAIGRVPAHPDIDSULTYSFERGENEISLVILNASTG 1080  
 Db 1021 PVUGNPEFILLNNYTTNRSSEFPGGAIGRVPAHPDIDSULTYSFERGENEISLVILNASTG 1080  
 Qy 1081 ELKLSRLDNNRPLEAISMVLSVDSGVASTAQCALRTIITDEMLTHSITLREDSMPER 1140  
 Db 1081 ELKLSRLDNNRPLEAISMVLSVDSGVASTAQCALRTIITDEMLTHSITLREDSMPER 1140  
 Qy 1141 FLSPBLGLFTQAAVAATLAPPDHVNVNORJDTDAPGHLINLVSLVSGPFGGCGPFL 1200  
 Db 1141 FLSPBLGLFTQAAVAATLAPPDHVNVNORJDTDAPGHLINLVSLVSGPFGGCGPFL 1200  
 Qy 1201 PPSDLOERLYLNSGLTLAISAGRYLPPDDNICLAREPENTMRCVSVLRFSSAPFLTSSS 1260  
 Db 1201 PPSDLOERLYLNSGLTLAISAGRYLPPDDNICLAREPENTMRCVSVLRFSSAPFLTSSS 1260  
 Qy 1261 VLFRPIHPVGLRCRCPGFTGVDCETEVDLCYSRPGPHGRCSRREGVTCCLCRDGYTG 1320  
 Db 1261 VLFRPIHPVGLRCRCPGFTGVDCETEVDLCYSRPGPHGRCSRREGVTCCLCRDGYTG 1320

Qy 1321 EHCEVSARSGRCTPGVCKNGKTCVNLIVGFKCDPSPGDEKPYCOVTTSPFAHSFITP 1380  
 Db 1321 EHCEVSARSGRCTPGVCKNGKTCVNLIVGFKCDPSPGDEKPYCOVTTSPFAHSFITP 1380  
 Qy 1381 RGLRORPHFTLALSPATKERDGLLLYNGRPNRKHDPVLALEVOYLTFSAGESSTTVS 1440  
 Db 1381 RGLRORPHFTLALSPATKERDGLLLYNGRPNRKHDPVLALEVOYLTFSAGESSTTVS 1440  
 Qy 1441 PFVPGVSDQWHTVQVLTQYNNKPLLGOTGLPOGSSQKXAVVTVVDCGDGVALARFSGVLG 1500  
 Db 1441 PFVPGVSDQWHTVQVLTQYNNKPLLGOTGLPOGSSQKXAVVTVVDCGDGVALARFSGVLG 1500  
 Qy 1501 NYSCAAQGTQSGSKSLDLTGPILLGAVPDLPSFPVNRKQFVGCNRNLTQVDSRHIDMAD 1560  
 Db 1501 NYSCAAQGTQSGSKSLDLTGPILLGAVPDLPSFPVNRKQFVGCNRNLTQVDSRHIDMAD 1560  
 Qy 1561 FIANNGTVPCCPKKQVCDNSTCHNGGTCVQMDAPSCBCEPLGFGKSCAQEMANPOHFL 1620  
 Db 1561 FIANNGTVPCCPKKQVCDNSTCHNGGTCVQMDAPSCBCEPLGFGKSCAQEMANPOHFL 1620  
 Qy 1621 GSSLVAMHGLSLPISQPWYLSIMFRTROADGVTLQITGRSTITLQLEBGMVLSVEGT 1680  
 Db 1621 GSSLVAMHGLSLPISQPWYLSIMFRTROADGVTLQITGRSTITLQLEBGMVLSVEGT 1680  
 Qy 1681 GLQASSLRLEPGRANDGDMHHAQLALGASGPGHALLSFYQGOARAGNLGPRHGLHLS 1740  
 Db 1681 GLQASSLRLEPGRANDGDMHHAQLALGASGPGHALLSFYQGOARAGNLGPRHGLHLS 1740  
 Qy 1741 NITVGGI PGPAAGVANGFRGCIQGVAVSPTPBGVNSLDPHSGESINVEGCSLDPDCSN 1800  
 Db 1741 NITVGGI PGPAAGVANGFRGCIQGVAVSPTPBGVNSLDPHSGESINVEGCSLDPDCSN 1800  
 Qy 1801 PCPANSYCSNDMDVSCSCDPGYGDNCTNVCIDLANCEHOSVCTRPSAPHGTCCCPBN 1860  
 Db 1801 PCPANSYCSNDMDVSCSCDPGYGDNCTNVCIDLANCEHOSVCTRPSAPHGTCCCPBN 1860  
 Qy 1861 YLGPYCETRIIDOPCPRGMMGHPTGCPNCNCDVSKGPPDCNKTSGECHKENHRRPGSPT 1920  
 Db 1861 YLGPYCETRIIDOPCPRGMMGHPTGCPNCNCDVSKGPPDCNKTSGECHKENHRRPGSPT 1920  
 Qy 1921 CLLDCCTPPGSGLSRVCDPEBQCPCKRGVYIGRCQRCNDNPAFVTTNGCBVATDSCPRAL 1980  
 Db 1921 CLLDCCTPPGSGLSRVCDPEBQCPCKRGVYIGRCQRCNDNPAFVTTNGCBVATDSCPRAL 1980  
 Qy 1981 EAGIMPRTRFGLPAAAPCPKGSFGTAVRHCEHRGMLPNNLFCNCTSIFFSELKGFABRL 2040  
 Db 1981 EAGIMPRTRFGLPAAAPCPKGSFGTAVRHCEHRGMLPNNLFCNCTSIFFSELKGFABRL 2040  
 Qy 2041 QRNESGLDSGRSQQLALLRNATOHYAGYFGSDVYKAYOLATRLAHBSTORGFGLSATQ 2100  
 Db 2041 QRNESGLDSGRSQQLALLRNATOHYAGYFGSDVYKAYOLATRLAHBSTORGFGLSATQ 2100  
 Qy 2101 DVHFTENLLRVGSALLDTANKRHWELIOQTBEGTAMLQHYEAYASALAKQNNRHLYLSPF 2160  
 Db 2101 DVHFTENLLRVGSALLDTANKRHWELIOQTBEGTAMLQHYEAYASALAKQNNRHLYLSPF 2160  
 Qy 2161 TIVTPNIVSVVRLDKGNFAGAKLPRYEALRGSEOPDLETYVILPESVRETPPVVRPAG 2220  
 Db 2161 TIVTPNIVSVVRLDKGNFAGAKLPRYEALRGSEOPDLETYVILPESVRETPPVVRPAG 2220  
 Qy 2221 PGBAQBEEELARRORRHEPLSQGRAVASVYIYRTLAGLLPHNYDPDKSLRVKRPDIINT 2280  
 Db 2221 PGBAQBEEELARRORRHEPLSQGRAVASVYIYRTLAGLLPHNYDPDKSLRVKRPDIINT 2280  
 Qy 2281 PVVSTISVHDEBELPALDKPVTVOFRLLETERKTPICVFNHSHLTVSGTGSMSRAGE 2340  
 Db 2281 PVVSTISVHDEBELPALDKPVTVOFRLLETERKTPICVFNHSHLTVSGTGSMSRAGE 2340  
 Qy 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFPFLTL 2400  
 Db 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFPFLTL 2400

QY 2401 LKILNSNGGIRRNLTALGLAQLVFLGIGNADLPACTVIAIILHPLYICTFSWALLE 2460  
DB 2401 LKILNSNGGIRRNLTALGLAQLVFLGIGNADLPACTVIAIILHPLYICTFSWALLE 2460  
QY 2461 ALHLVYALVEVDVNTGMPFVYMLQWVPAFTTGLAVGLDPBGYKGNPFCWLSITDTLI 2520  
DB 2461 ALHLVYALVEVDVNTGMPFVYMLQWVPAFTTGLAVGLDPBGYKGNPFCWLSITDTLI 2520  
QY 2521 WSPFAGVAVAVSVSVLYTILAAASCAARQCFEKKGPVSGLOPFAVILLISATWILAL 2580  
DB 2521 WSPFAGVAVAVSVSVLYTILAAASCAARQCFEKKGPVSGLOPFAVILLISATWILAL 2580  
QY 2581 LSVNSTLTLLPHLIFATCNCCIQGFPIFLSYVTSKEYRKALKIACSKRPSPDPALTTKSTL 2640  
DB 2581 LSVNSTLTLLPHLIFATCNCCIQGFPIFLSYVTSKEYRKALKIACSKRPSPDPALTTKSTL 2640  
QY 2641 TSSYNCPSPYADGRLYOPYGDSDAGSLHSTSRSGKSPSYTPLLREBSALNPGQGPFG 2700  
DB 2641 TSSYNCPSPYADGRLYOPYGDSDAGSLHSTSRSGKSPSYTPLLREBSALNPGQGPFG 2700  
QY 2701 DPGSLFLBQDQDHDPTDSDLSLEDDQSGSYASTHSSDEEEREEREAEAFPGEG 2760  
DB 2701 DPGSLFLBQDQDHDPTDSDLSLEDDQSGSYASTHSSDEEEREEREAEAFPGEG 2760  
QY 2761 WPSLCPGAEPLPLHSTPDGPGPGKAPWPGDFTTAKSSGNGAPDEERLRENGDALSR 2820  
DB 2761 WPSLCPGAEPLPLHSTPDGPGPGKAPWPGDFTTAKSSGNGAPDEERLRENGDALSR 2820  
QY 2821 EBSLGLPSSAOPHGIILKKKCLPTISEKSLRLPLEQCTGSSGSSASBESRGGPP 2880  
DB 2821 EBSLGLPSSAOPHGIILKKKCLPTISEKSLRLPLEQCTGSSGSSASBESRGGPP 2880  
QY 2881 RPPRQSLQEOULNGVWPIAMSIKAGTVDEDSGSEFLPFNFILH 2923  
DB 2881 RPPRQSLQEOULNGVWPIAMSIKAGTVDEDSGSEFLPFNFILH 2923

RESULT 2  
US-09-916-849A-3  
; Sequence 3, Application US/0916849A  
; Publication No. US20030086934A1  
; GENERAL INFORMATION:  
; APPLICANT: Boctein, et al.  
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents  
; TITLE OF INVENTION: Uses thereof  
; FILE REFERENCE: 2002850-0024  
; CURRENT APPLICATION NUMBER: US/09/916,849A  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2923  
; TYPE: PRF  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Cadherin EGF  
; OTHER INFORMATION: IAG Seven Pass G-Type Receptor 2  
US-09-916-849A-3

Query Match 100.0%; Score 2923; DB 11; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLIGDQVPCRSLSGRSGSGACA PMGMLCPS 60  
DB 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLIGDQVPCRSLSGRSGSGACA PMGMLCPS 60  
QY 61 SASNLMLYTSRCRDAGTELTHLVPHHDSLRVWCPESEAHIPLPAPAECCPMSCRLIGIG 120  
DB 61 SASNLMLYTSRCRDAGTELTHLVPHHDSLRVWCPESEAHIPLPAPAECCPMSCRLIGIG 120  
QY 121 GHLSPGKLTTPBEHPCAKAPRLRCQSCKLAQAPGIRAGERSBESLGGRRRRNVTAPO 180  
DB 121 GHLSPGKLTTPBEHPCAKAPRLRCQSCKLAQAPGIRAGERSBESLGGRRRRNVTAPO 180

DB 121 GHLSPGKLTTPBEHPCAKAPRLRCQSCKLAQAPGIRAGERSBESLGGRRRRNVTAPO 180  
QY 181 FQPPSYQATVPENOPAGTPVASTRAIDPBGAGRLTYMDALPDSRSNOFSLDPVTA 240  
DB 181 FQPPSYQATVPENOPAGTPVASTRAIDPBGAGRLTYMDALPDSRSNOFSLDPVTA 240  
QY 241 VTTAEELRETKSTVFRVTAQDHGMPRRSALATLTITLVTDNHDVPVEQOEYKESLRE 300  
DB 241 VTTAEELRETKSTVFRVTAQDHGMPRRSALATLTITLVTDNHDVPVEQOEYKESLRE 300  
QY 301 NIEVEGYVLTARATGDAPPPANITLRLLEGSGSPSEFEIDPBGVTRTGPDVREY 360  
DB 301 NIEVEGYVLTARATGDAPPPANITLRLLEGSGSGSPSEFEIDPBGVTRTGPDVREY 360  
QY 361 ESYQOLTEASDQGRDPGPRSTAAVFLSVEDNDNAPQSEKRYVQVQVEDVTPGAPVR 420  
DB 361 ESYQOLTEASDQGRDPGPRSTAAVFLSVEDNDNAPQSEKRYVQVQVEDVTPGAPVR 420  
QY 421 VTASDRDKSNVAVHYSIMSGNARQFYLDAGTALDVVSPLDYETTKETTLRVADQDG 480  
DB 421 VTASDRDKSNVAVHYSIMSGNARQFYLDAGTALDVVSPLDYETTKETTLRVADQDG 480  
QY 481 RPPLSNVSGLVTVQVLDINDAPIFVSTPFOATVLESVPLGVLVHQAIDADAGDNAL 540  
DB 481 RPPLSNVSGLVTVQVLDINDAPIFVSTPFOATVLESVPLGVLVHQAIDADAGDNAL 540  
QY 541 EYRLAGVGHDPFTLNNGTGMSVAEILDRBEVDYFSGEVARDHGTPLTASASVTV 600  
DB 541 EYRLAGVGHDPFTLNNGTGMSVAEILDRBEVDYFSGEVARDHGTPLTASASVTV 600  
QY 601 LDVNDNPPFTQPEXTVRLNEDAAVGTSVVTSADVDRSHSVTYQITSGNTRNFSITS 660  
DB 601 LDVNDNPPFTQPEXTVRLNEDAAVGTSVVTSADVDRSHSVTYQITSGNTRNFSITS 660  
QY 661 QSGGGLVSLALPLDYKLRQVYLAATASDGTRODTRQIYVNTDANTHRPVQSSHYTN 720  
DB 661 QSGGGLVSLALPLDYKLRQVYLAATASDGTRODTRQIYVNTDANTHRPVQSSHYTN 720  
QY 721 VNEBDRPAGTIVLISATBEDTGENARITYFMEDSIPQIRIDADTAVTTOALDEDOVS 780  
DB 721 VNEBDRPAGTIVLISATBEDTGENARITYFMEDSIPQIRIDADTAVTTOALDEDOVS 780  
QY 781 YTLAATARDNGIPQKSDTTYLEILVNDVNDNAPQFLRDSYQSVYEDVPPFTSVLQISAT 840  
DB 781 YTLAATARDNGIPQKSDTTYLEILVNDVNDNAPQFLRDSYQSVYEDVPPFTSVLQISAT 840  
QY 841 DRDSGLNRRVPTFGQGDGDDPFVSESTSGVTRLRDLRENVAAQYVLRAYAVDKGMP 900  
DB 841 DRDSGLNRRVPTFGQGDGDDPFVSESTSGVTRLRDLRENVAAQYVLRAYAVDKGMP 900  
QY 901 ARTPMETVTVLADVNDNPPVFEQDEDFVFEENSPIGLAVARVATDDEGTAQIMYQI 960  
DB 901 ARTPMETVTVLADVNDNPPVFEQDEDFVFEENSPIGLAVARVATDDEGTAQIMYQI 960  
QY 961 VEGNIPEVFOLDIFSGELTALVDLYEDRPREVYVLOATSAPLVSRATVHVALLDRNDP 1020  
DB 961 VEGNIPEVFOLDIFSGELTALVDLYEDRPREVYVLOATSAPLVSRATVHVALLDRNDP 1020  
QY 1021 PVLGNFELFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFERGMBLSVYLANASTG 1080  
DB 1021 PVLGNFELFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFERGMBLSVYLANASTG 1080  
QY 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTAQCALRVITIDEMKTHSITTLRLDEMSPER 1140  
DB 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTAQCALRVITIDEMKTHSITTLRLDEMSPER 1140  
QY 1141 FLSPPLGFLIOVAATLATPPDHVVVENVORDTAPGGHILNVSLVSCQPPGGGPPFL 1200  
DB 1141 FLSPPLGFLIOVAATLATPPDHVVVENVORDTAPGGHILNVSLVSCQPPGGGPPFL 1200  
QY 1201 PSEDLQERLYINRSILTAISQORVLPDDNITCLAREPCENYKCVSVLRFDSAPPIASS 1260  
DB 1201 PSEDLQERLYINRSILTAISQORVLPDDNITCLAREPCENYKCVSVLRFDSAPPIASS 1260

QY 1261 VLEFPIHPVGLACRCPPGFTGPDYCEYVLCYSRPGCPHRCRSGRGVYLCRDGYG 1320  
 DB 1261 VLFPIHPVGLACRCPPGFTGPDYCEYVLCYSRPGCPHRCRSGRGVYLCRDGYG 1320  
 QY 1321 EHCEVASRSGRCPGVCKNGKTCVNLVSGFKDCDCEDEKPYCQVTRSPHSPITE 1380  
 DB 1321 EHCEVASRSGRCPGVCKNGKTCVNLVSGFKDCDCEDEKPYCQVTRSPHSPITE 1380  
 QY 1381 RGRORPHFTLALSPATKEDGILLNGRFNEKHDPALEIOBOVOLTSSABESTTYS 1440  
 DB 1381 RGRORPHFTLALSPATKEDGILLNGRFNEKHDPALEIOBOVOLTSSABESTTYS 1440  
 QY 1441 PFPVGVSDQGMHTVOLKTYNKEPLAGOTGLPGSPBOKVAVTVGDCDVTALRFGSVLG 1500  
 DB 1441 PFPVGVSDQGMHTVOLKTYNKEPLAGOTGLPGSPBOKVAVTVGDCDVTALRFGSVLG 1500  
 QY 1501 NYSCAAQGTGGSKSIDLTGPIILGVPDLPSFPVRMRQFVGCNRNLTQVDSRHIDMAD 1560  
 DB 1501 NYSCAAQGTGGSKSIDLTGPIILGVPDLPSFPVRMRQFVGCNRNLTQVDSRHIDMAD 1560  
 QY 1561 FIANNQVPGCPAKKNCVDSNTCHNGGTGVNMDASCECPILRFGKSCAQMENAPQHL 1620  
 DB 1561 FIANNQVPGCPAKKNCVDSNTCHNGGTGVNMDASCECPILRFGKSCAQMENAPQHL 1620  
 QY 1621 GSSLVAMHGLSLPIQPMWYLSMFRTRQADGVLLQAITRGRSTTLQLRGHWLSEGT 1680  
 DB 1621 GSSLVAMHGLSLPIQPMWYLSMFRTRQADGVLLQAITRGRSTTLQLRGHWLSEGT 1680  
 QY 1681 GLOASSILREPRANDMDWHHAOLALGASGPGHAILSPDYGOORAGNIGPRLGLHLIS 1740  
 DB 1681 GLOASSILREPRANDMDWHHAOLALGASGPGHAILSPDYGOORAGNIGPRLGLHLIS 1740  
 QY 1741 NITVGGIPGAGVANGFRGCLQGVRYSDTPEGVNSIDPSHGESINVEGCSLPDPDSN 1800  
 DB 1741 NITVGGIPGAGVANGFRGCLQGVRYSDTPEGVNSIDPSHGESINVEGCSLPDPDSN 1800  
 QY 1801 PCPANSYCSNDMDSYSCSDPGYGDNCNTVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 DB 1801 PCPANSYCSNDMDSYSCSDPGYGDNCNTVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 QY 1861 YLGPYCESTRIDOPCFRGMWGHPTGCPNCVSKGFPDCKNTSGECKENHTRPPGSP 1920  
 DB 1861 YLGPYCESTRIDOPCFRGMWGHPTGCPNCVSKGFPDCKNTSGECKENHTRPPGSP 1920  
 QY 1921 CLICDCYPTGSLSRVCDPEDGQCPCKPGVIGROCDRCNDPFAVYTNNGCEVNTDSCRAI 1980  
 DB 1921 CLICDCYPTGSLSRVCDPEDGQCPCKPGVIGROCDRCNDPFAVYTNNGCEVNTDSCRAI 1980  
 QY 1981 EAGIWPRTFGLPAAAPCEKSGFPTAVRHCDEHGMPLNLFUNCTSITPSELKGFAPRL 2040  
 DB 1981 EAGIWPRTFGLPAAAPCEKSGFPTAVRHCDEHGMPLNLFUNCTSITPSELKGFAPRL 2040  
 QY 2041 QNRBSGIDSRSQLALLRNATOTHTAGTGS DYKAVAYOLATRLANESYORGGISATQ 2100  
 DB 2041 QNRBSGIDSRSQLALLRNATOTHTAGTGS DYKAVAYOLATRLANESYORGGISATQ 2100  
 QY 2101 DVHFTETLAVGSALLPTANKRHMLIQOTEGGTAMLQHYEYASALQNMHTYISPF 2160  
 DB 2101 DVHFTETLAVGSALLPTANKRHMLIQOTEGGTAMLQHYEYASALQNMHTYISPF 2160  
 QY 2161 TIVTPNIVISVRLDKGNFAGAKLPRYEALRGEQPPDLFTTVILPESVFRETPEVVPAG 2220  
 DB 2161 TIVTPNIVISVRLDKGNFAGAKLPRYEALRGEQPPDLFTTVILPESVFRETPEVVPAG 2220  
 QY 2221 PGEAQBEBELARORRHEPBLISQGBAVASVITRTLAGLPHNYDPPDRSRVPRPILNT 2280  
 DB 2221 PGEAQBEBELARORRHEPBLISQGBAVASVITRTLAGLPHNYDPPDRSRVPRPILNT 2280  
 QY 2281 PVSISVHDBELLPRALDKPVTVQFRLTEERTKEICVFMNHSILVSGTGMSARGCE 2340  
 DB 2281 PVSISVHDBELLPRALDKPVTVQFRLTEERTKEICVFMNHSILVSGTGMSARGCE 2340

QY 2341 VVFNESHVSQCNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLLALHTFEFLT 2400  
 DB 2341 VVFNESHVSQCNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLLALHTFEFLT 2400  
 QY 2401 LRILSNQHGIRRNITLALGLAQLVFLGINADLPACTVIAIILHFTYLCTFSWALLE 2460  
 DB 2401 LRILSNQHGIRRNITLALGLAQLVFLGINADLPACTVIAIILHFTYLCTFSWALLE 2460  
 QY 2461 ALHIVRALTEVDVMTGPMRFYMLGMGVPARTITGLAVGLDREGVNPFCMLISTDTLI 2520  
 DB 2461 ALHIVRALTEVDVMTGPMRFYMLGMGVPARTITGLAVGLDREGVNPFCMLISTDTLI 2520  
 QY 2521 WSPAGVAVASVSYFLYTLAARASCAQROGFEKKGPVSGLOPSPAVILLISATWLLAL 2580  
 DB 2521 WSPAGVAVASVSYFLYTLAARASCAQROGFEKKGPVSGLOPSPAVILLISATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLPATCNCIQPPTFLSYVTSKRYKALKACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLPATCNCIQPPTFLSYVTSKRYKALKACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKSQPSYIPLLRBSALNPGQGPGLG 2700  
 DB 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKSQPSYIPLLRBSALNPGQGPGLG 2700  
 QY 2701 DPGSLFLEGQDOHDPDTSDDLILKDDQSGSYASTHSDBEEBEEBEEBEEBAAFPGEQ 2760  
 DB 2701 DPGSLFLEGQDOHDPDTSDDLILKDDQSGSYASTHSDBEEBEEBEEBEEBAAFPGEQ 2760  
 QY 2761 WDSLAPGABRLPLHSTPDGPGPKAPWPDGFTTAKSSGNAPBERLRENDALSR 2820  
 DB 2761 WDSLAPGABRLPLHSTPDGPGPKAPWPDGFTTAKSSGNAPBERLRENDALSR 2820  
 QY 2821 EBSLGLPGSSAOPPHGILKXKCLPTISEKSLRLPLEQCSSSRSGSSRSGGPP 2880  
 DB 2821 EBSLGLPGSSAOPPHGILKXKCLPTISEKSLRLPLEQCSSSRSGSSRSGGPP 2880  
 QY 2881 RPPRQSLQEQLVGWPVIMSIIKAGTVDESSGSEFLFENFLH 2923  
 DB 2881 RPPRQSLQEQLVGWPVIMSIIKAGTVDESSGSEFLFENFLH 2923

RESULT 3  
 US-10-174-677-29  
 ; Sequence 29, Application US/10174677  
 ; Publication No. US20030190704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xie, Ting  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR  
 ; FILE REFERENCE: 40716(1P-012)  
 ; CURRENT APPLICATION NUMBER: US/10/174,677  
 ; CURRENT FILING DATE: 2002-06-19  
 ; NUMBER OF SEQ ID NOS: 117  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-174-677-29

Query Match 100.0%; Score 2923; DB 12; Length 2923;  
 Best Local Similarity 100.0%; Fred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSPATGVLPPPPPLILILILILPPPLIGQVGCRLASRGSSGACAPMGWLCPS 60  
 DB 1 MSPATGVLPPPPPLILILILILPPPLIGQVGCRLASRGSSGACAPMGWLCPS 60  
 QY 61 SASNLMLYTSRCKDAGTELTGHLVPHHDLRYWCPSBSAHILPPAPBEGCPMSCLLGTG 120  
 DB 61 SASNLMLYTSRCKDAGTELTGHLVPHHDLRYWCPSBSAHILPPAPBEGCPMSCLLGTG 120  
 QY 121 GHLSPGKTLTPEHPCLAPRLRCQSCGLAQPGRAGERSPEBSLGGRRKNVTAPO 180  
 DB 121 GHLSPGKTLTPEHPCLAPRLRCQSCGLAQPGRAGERSPEBSLGGRRKNVTAPO 180

Db 121 GHLISQKCLTLEBHPCLKAPRLKQSCKLAQAPGIRAGERSPEBSLGRRRRNVNTAQ 180  
 Qy 181 FQPSYQATVPENQAGTVPASALRAIDPDEBAGLEETMDLPDSRSNQFSLDPVTGA 240  
 Db 181 FQPSYQATVPENQAGTVPASALRAIDPDEBAGLEETMDLPDSRSNQFSLDPVTGA 240  
 Qy 241 VTTAEELDRETKSTIVFRVTAODHGMRRSALATLTILVTDTNDHPVEEQEYKESLRE 300  
 Db 241 VTTAEELDRETKSTIVFRVTAODHGMRRSALATLTILVTDTNDHPVEEQEYKESLRE 300  
 Qy 301 NLEVEYEVLTARAATGDA.PPNNANILYRLLESGSGSPSEVFEIDPDSGVITRTGPDREBY 360  
 Db 301 NLEVEYEVLTARAATGDA.PPNNANILYRLLESGSGSPSEVFEIDPDSGVITRTGPDREBY 360  
 Qy 361 ESYQULVEASDQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVQVREEDVTPCAPYLR 420  
 Db 361 ESYQULVEASDQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVQVREEDVTPCAPYLR 420  
 Qy 421 VTTASRDVGSNAVHTYSTMSGNARQFYLDAGTALDVVSPLDYETTKYTLRVARADQG 480  
 Db 421 VTTASRDVGSNAVHTYSTMSGNARQFYLDAGTALDVVSPLDYETTKYTLRVARADQG 480  
 Qy 481 RPLSNVSGLVTVQVLIDINDNAPIFVSTPFOATVLESVPLGYLHVOAIDADADNARL 540  
 Db 481 RPLSNVSGLVTVQVLIDINDNAPIFVSTPFOATVLESVPLGYLHVOAIDADADNARL 540  
 Qy 541 EYRLAGVHDEPFTINNGTGMISVAAELDREVDYFSFGVEARDHGTALTSASVSTV 600  
 Db 541 EYRLAGVHDEPFTINNGTGMISVAAELDREVDYFSFGVEARDHGTALTSASVSTV 600  
 Qy 601 LDVNDNFTPOPEYTVRLNEDAVGTSVTVSAVDRAHVSITVQITSGNTRNFSITS 660  
 Db 601 LDVNDNFTPOPEYTVRLNEDAVGTSVTVSAVDRAHVSITVQITSGNTRNFSITS 660  
 Qy 661 OSGGGLVSLALPLDYKLEROYVLAATASDGTRODPAQIVNNTANTHRPFOSSHHTVN 720  
 Db 661 OSGGGLVSLALPLDYKLEROYVLAATASDGTRODPAQIVNNTANTHRPFOSSHHTVN 720  
 Qy 721 VNEBRPACTTVVLSATDEDTGENARITYFMEDSIPOFRIDADTGAVTTQAELEDPOVS 780  
 Db 721 VNEBRPACTTVVLSATDEDTGENARITYFMEDSIPOFRIDADTGAVTTQAELEDPOVS 780  
 Qy 781 YTLAITARDNGI.POKSDITTYLEILVNDVNDNAPOFLRDSYQSGYEDVPFTSVLQISAT 840  
 Db 781 YTLAITARDNGI.POKSDITTYLEILVNDVNDNAPOFLRDSYQSGYEDVPFTSVLQISAT 840  
 Qy 841 DRDSGLNRRVFTYTOGGDGDGDDPFTVESTSGIVRTLRLDRBNVAQVYLRAVADKGMPP 900  
 Db 841 DRDSGLNRRVFTYTOGGDGDGDDPFTVESTSGIVRTLRLDRBNVAQVYLRAVADKGMPP 900  
 Qy 901 ARTPEVTVTVLDVNDNPFVEQDEDFDVFEENSPIGLAVARVTAATDDEGTNAQIMYQI 960  
 Db 901 ARTPEVTVTVLDVNDNPFVEQDEDFDVFEENSPIGLAVARVTAATDDEGTNAQIMYQI 960  
 Qy 961 VEGNIPEVFOULDISGELTALVDLDYEDRPEVIVIQATSAPLVSRATVHARLLDRNDP 1020  
 Db 961 VEGNIPEVFOULDISGELTALVDLDYEDRPEVIVIQATSAPLVSRATVHARLLDRNDP 1020  
 Qy 1021 PVLDNFEELFNNVYTNBSSSPRGALIGRVPAHDPDISLTYSPERGHELSLVILNASTG 1080  
 Db 1021 PVLDNFEELFNNVYTNBSSSPRGALIGRVPAHDPDISLTYSPERGHELSLVILNASTG 1080  
 Qy 1081 ELKLSRALDNNRPLEAISMVSVSDGHSVTACALRVITIDEMULTHSITLRLDEMSPER 1140  
 Db 1081 ELKLSRALDNNRPLEAISMVSVSDGHSVTACALRVITIDEMULTHSITLRLDEMSPER 1140  
 Qy 1141 FLUSPLGLFIOAVAAATLTPPDHVVVNVQDTPDAPGGHIIINVSISVGQPPGPGGPPFL 1200  
 Db 1141 FLUSPLGLFIOAVAAATLTPPDHVVVNVQDTPDAPGGHIIINVSISVGQPPGPGGPPFL 1200  
 Qy 1201 PSEDLQEBLYNBSLTLTAISAORVLP.PPDNCLAREPCENMR.CVSVLRFDSAPFIASSS 1260  
 Db 1201 PSEDLQEBLYNBSLTLTAISAORVLP.PPDNCLAREPCENMR.CVSVLRFDSAPFIASSS 1260

Qy 1261 VLFPRIFHEVGGI.RCPCPGFTGDYCETEVND.CYSRNPQCPHGRCRREBGGYTCLACDGYTG 1320  
 Db 1261 VLFPRIFHEVGGI.RCPCPGFTGDYCETEVND.CYSRNPQCPHGRCRREBGGYTCLACDGYTG 1320  
 Qy 1321 EHCEVSASRGCTPGVCNKGTCVNLVGGFKCDPCPSGDFEKPYCQVTTSPAPHSITTF 1380  
 Db 1321 EHCEVSASRGCTPGVCNKGTCVNLVGGFKCDPCPSGDFEKPYCQVTTSPAPHSITTF 1380  
 Qy 1381 RGLRORFHFITALSPATERPDGLLXNGRFBKEDFVALFVIOBYOULTPSAGSSTTVS 1440  
 Db 1381 RGLRORFHFITALSPATERPDGLLXNGRFBKEDFVALFVIOBYOULTPSAGSSTTVS 1440  
 Qy 1441 PFVFGVSDQGMHTVQLYKYRKPLLGGTGLPQGSBQVAVVTVYDGCCTGVALRSGSYLG 1500  
 Db 1441 PFVFGVSDQGMHTVQLYKYRKPLLGGTGLPQGSBQVAVVTVYDGCCTGVALRSGSYLG 1500  
 Qy 1501 NYSQAQGTQGSKKSIDLTGPLLGGVPLD.PBSFPVARMQPFVGMRLQVDSRHIDMAD 1560  
 Db 1501 NYSQAQGTQGSKKSIDLTGPLLGGVPLD.PBSFPVARMQPFVGMRLQVDSRHIDMAD 1560  
 Qy 1561 FIANNGTVPGRPAKNVCDSTNCHNGTCVNOAMPASCECEPLGFGKSCAQEMANPOHFL 1620  
 Db 1561 FIANNGTVPGRPAKNVCDSTNCHNGTCVNOAMPASCECEPLGFGKSCAQEMANPOHFL 1620  
 Qy 1621 GSSLVAMHGLSLPI.SOPPWYLSIMFRTROADGVLLQAITRGSTITTLQLEBGMVLSYBGT 1680  
 Db 1621 GSSLVAMHGLSLPI.SOPPWYLSIMFRTROADGVLLQAITRGSTITTLQLEBGMVLSYBGT 1680  
 Qy 1681 GLQASLRLBEPGRANDGDWHHAQALASGGPGHAILSFYDQGRABGNLGRLLHGLHS 1740  
 Db 1681 GLQASLRLBEPGRANDGDWHHAQALASGGPGHAILSFYDQGRABGNLGRLLHGLHS 1740  
 Qy 1741 NITVGGI.PGPAAGVARGPRGCLQGVRSVDPPEGVNSLDPHSGESINVEQGSLLDPCCSN 1800  
 Db 1741 NITVGGI.PGPAAGVARGPRGCLQGVRSVDPPEGVNSLDPHSGESINVEQGSLLDPCCSN 1800  
 Qy 1801 PCPANSYCSNDMDPSYSCSPGYGDNCTVNCIDNPCEHOSVCRKPSAPRGYCECCEPN 1860  
 Db 1801 PCPANSYCSNDMDPSYSCSPGYGDNCTVNCIDNPCEHOSVCRKPSAPRGYCECCEPN 1860  
 Qy 1861 YLGPYCESTRIDQCP.RPGMKGHPITGPCNCNVSKGFPDPCNKTSGECHKENHYRPPGSPPT 1920  
 Db 1861 YLGPYCESTRIDQCP.RPGMKGHPITGPCNCNVSKGFPDPCNKTSGECHKENHYRPPGSPPT 1920  
 Qy 1921 CLLCDYPTGSLASVCDPEBQGCCKPQVIGROCDRCNDNPAEYTTNGCEFNYSCPRAI 1980  
 Db 1921 CLLCDYPTGSLASVCDPEBQGCCKPQVIGROCDRCNDNPAEYTTNGCEFNYSCPRAI 1980  
 Qy 1981 BAGIMMPRTFRGLPAAAPCPKSGFATAVRHCDERHGM.LPPLFACTSITPSELKGFABRL 2040  
 Db 1981 BAGIMMPRTFRGLPAAAPCPKSGFATAVRHCDERHGM.LPPLFACTSITPSELKGFABRL 2040  
 Qy 2041 QRNESGLDSRSGQALILNNAOTHTAGYFGSDVKAAYOLATRLLAHSTORGFLSATQ 2100  
 Db 2041 QRNESGLDSRSGQALILNNAOTHTAGYFGSDVKAAYOLATRLLAHSTORGFLSATQ 2100  
 Qy 2101 DVHFTENLTVGSAILLDTANKRHMBELI.QOTBGTAMLLQHYEAAASALAQMRRTYLSPF 2160  
 Db 2101 DVHFTENLTVGSAILLDTANKRHMBELI.QOTBGTAMLLQHYEAAASALAQMRRTYLSPF 2160  
 Qy 2161 TIVTVPIVIVSVRLDKNPAAGAKLPRYEALRGEOPLLETTVILPESVFRET.PVVPAPAG 2220  
 Db 2161 TIVTVPIVIVSVRLDKNPAAGAKLPRYEALRGEOPLLETTVILPESVFRET.PVVPAPAG 2220  
 Qy 2221 PGEAQBEBELARRORRHELSQGEAVASVIIYRTLAGLLPHNTYDPDRSLRVPKPIINT 2280  
 Db 2221 PGEAQBEBELARRORRHELSQGEAVASVIIYRTLAGLLPHNTYDPDRSLRVPKPIINT 2280  
 Qy 2281 PVSISVHDBEELPRALDKPVTVQFRLLEBETKRI.CVPMNHSIIVSGTGSARGCE 2340  
 Db 2281 PVSISVHDBEELPRALDKPVTVQFRLLEBETKRI.CVPMNHSIIVSGTGSARGCE 2340



QY 2341 VVERNSHVSQCQNMHTSPAVLMDVSRRENGEILLPLKTLTYVALGVTLAALLTPPELTL 2400  
DB 2341 VVERNSHVSQCQNMHTSPAVLMDVSRRENGEILLPLKTLTYVALGVTLAALLTPPELTL 2400  
QY 2401 LRLTRNNGHGRNRLTRALGLAQLVPLLGINODLPACTVIAILLHFTLCTPSALLL 2460  
DB 2401 LRLTRNNGHGRNRLTRALGLAQLVPLLGINODLPACTVIAILLHFTLCTPSALLL 2460  
QY 2461 ALHLVYALTEVDVNTGPMRFYMLGWVPAFITGLAVGLDPGNGNDECMLSIYDTLI 2520  
DB 2461 ALHLVYALTEVDVNTGPMRFYMLGWVPAFITGLAVGLDPGNGNDECMLSIYDTLI 2520  
QY 2521 WSPAGVAFVAVSVSVLYTLAARASCAORQGFEGKGPVGLQPSFAVILLLSATWLLAL 2580  
DB 2521 WSPAGVAFVAVSVSVLYTLAARASCAORQGFEGKGPVGLQPSFAVILLLSATWLLAL 2580  
QY 2581 LSVNSDTLLPHYLPAFCNCIQGPFILSYVLSKEVRKALKACSRKPSDPAALTTKSTL 2640  
DB 2581 LSVNSDTLLPHYLPAFCNCIQGPFILSYVLSKEVRKALKACSRKPSDPAALTTKSTL 2640  
QY 2641 TSSVNCPSPYADGRLYQPYGDSAGSLHSTSRGSKQPSYIPFLIRBSALNPGQPPGLG 2700  
DB 2641 TSSVNCPSPYADGRLYQPYGDSAGSLHSTSRGSKQPSYIPFLIRBSALNPGQPPGLG 2700  
QY 2701 DPESLFLBEGDDQDHPDSDLSLEDDQSGSYASTHSSDSEEBEERBAAPFGEQG 2760  
DB 2701 DPESLFLBEGDDQDHPDSDLSLEDDQSGSYASTHSSDSEEBEERBAAPFGEQG 2760  
QY 2761 WDSLILPGARERLPLHSTPKDQGPGRKAPWPGTGTAKSSGNGAPBERLIRNGALSR 2820  
DB 2761 WDSLILPGARERLPLHSTPKDQGPGRKAPWPGTGTAKSSGNGAPBERLIRNGALSR 2820  
QY 2821 EGSIGLPLGSSAAPHKGLLKKCLPTISEKSLRLPLEQCTSSSGSSASBSRGAPPP 2880  
DB 2821 EGSIGLPLGSSAAPHKGLLKKCLPTISEKSLRLPLEQCTSSSGSSASBSRGAPPP 2880  
QY 2881 RPPRQSLQEQNLGVMPVAMSIRAGTYDEDSGSEPLFFNFLLH 2923  
DB 2881 RPPRQSLQEQNLGVMPVAMSIRAGTYDEDSGSEPLFFNFLLH 2923

RESULT 4  
US-10-120-801-53  
; Sequence 53, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Shinkete, Richard  
; APPLICANT: Padigaru, Muraidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Mehriban, Ruad  
; APPLICANT: Topper, James N  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT APPLICATION NUMBER: US/10/120, 801  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/285748  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/288334  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/291241

QY PRIOR FILING DATE: 2001-05-16  
DB PRIOR APPLICATION NUMBER: 60/322284  
QY PRIOR FILING DATE: 2001-09-14  
DB PRIOR APPLICATION NUMBER: 60/285609  
QY PRIOR FILING DATE: 2001-04-20  
DB NUMBER OF SEQ ID NOS: 155  
QY SOFTWARE: Patentin Ver. 2.1  
DB SEQ ID NO 53  
QY LENGTH: 2923  
DB TYPE: PRT  
QY ORGANISM: Drosophila melanogaster  
DB US-10-120-801-53

Query Match 100.0%; Score 2923; DB 12; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESPATGVPPLPPPPPLLLLLLLLLLPPPLLDQVPCBSLSRGSGSACAPMGWLCPS 60  
DB 1 MESPATGVPPLPPPPPLLLLLLLLLLPPPLLDQVPCBSLSRGSGSACAPMGWLCPS 60  
QY 61 SASNMLVTSRCRDAGTGLVPHHGLRVWCPESEAHPLPAPBSCPMSCRLGIG 120  
DB 61 SASNMLVTSRCRDAGTGLVPHHGLRVWCPESEAHPLPAPBSCPMSCRLGIG 120  
QY 121 GHLSFGKLTLPBEPHCLKAPRLRCQSCKLAQAPGLRAGERSPEBSLGGRRRNNVTAPO 180  
DB 121 GHLSFGKLTLPBEPHCLKAPRLRCQSCKLAQAPGLRAGERSPEBSLGGRRRNNVTAPO 180  
QY 181 FQPPSYOATVPENQAGTPVASLRAIDPDEGAGRLBETMDLFPDRSNQFSLDPVICA 240  
DB 181 FQPPSYOATVPENQAGTPVASLRAIDPDEGAGRLBETMDLFPDRSNQFSLDPVICA 240  
QY 241 VTTARELBREKTSHTVFVTADHGMPPRSALATLTITVTDNHDPVEOQETRESLRE 300  
DB 241 VTTARELBREKTSHTVFVTADHGMPPRSALATLTITVTDNHDPVEOQETRESLRE 300  
QY 301 NLEVGVEVLTVRATDGDAPPNANILYRLBGSQSPSEVFEIDPRSGVLRTRGPVDRBEV 360  
DB 301 NLEVGVEVLTVRATDGDAPPNANILYRLBGSQSPSEVFEIDPRSGVLRTRGPVDRBEV 360  
QY 361 ESYQUTVERASDQGRPGRRSTTAANVFLSVEDNDNDNAPOPSERKRYVQVREDVTPGAPVLR 420  
DB 361 ESYQUTVERASDQGRPGRRSTTAANVFLSVEDNDNDNAPOPSERKRYVQVREDVTPGAPVLR 420  
QY 421 VTASDRDKSNNAVHYSIMSGNARQFYLDAGTGLDVVSPLDYETTKETTLRVARQDGG 480  
DB 421 VTASDRDKSNNAVHYSIMSGNARQFYLDAGTGLDVVSPLDYETTKETTLRVARQDGG 480  
QY 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPPQATVYLSVPLGYLVHQAIDADAGDARL 540  
DB 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPPQATVYLSVPLGYLVHQAIDADAGDARL 540  
QY 541 EYRLAGVGHDPFTINNGTGMISVAEILDREVDVYSQVBARHDKGTPLTASASVTV 600  
DB 541 EYRLAGVGHDPFTINNGTGMISVAEILDREVDVYSQVBARHDKGTPLTASASVTV 600  
QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYSAVDRASHVITYQITSGNTRNFSITS 660  
DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYSAVDRASHVITYQITSGNTRNFSITS 660  
QY 661 QGGGGLVSLALPLDYKLRQVYLAVTASDGTQRQTAQIVVNTDANTHRPVQSSHYYTN 720  
DB 661 QGGGGLVSLALPLDYKLRQVYLAVTASDGTQRQTAQIVVNTDANTHRPVQSSHYYTN 720  
QY 721 VNEDRPAGTIVVLSATBEDGSENNRITYFEMDSIPORRIDDTAGVTAQALDVEDQVS 780  
DB 721 VNEDRPAGTIVVLSATBEDGSENNRITYFEMDSIPORRIDDTAGVTAQALDVEDQVS 780  
QY 781 YTLAATTARDNGIPQSDTYYLEILVNDVNDNAPQGLRDSYQGSYVEDVPFTSVQISAT 840  
DB 781 YTLAATTARDNGIPQSDTYYLEILVNDVNDNAPQGLRDSYQGSYVEDVPFTSVQISAT 840

QY 841 DRDGLNRRVYTTTGGGDDGDDPIVESTSGIVRTLRRLDRBNVAQYTLAAYADKMP 900  
 DB 841 DRDGLNRRVYTTTGGGDDGDDPIVESTSGIVRTLRRLDRBNVAQYTLAAYADKMP 900  
 QY 901 ARTPEMTVTVLADVNDNPFVEFODEDFVEENSTIGLAVARVTAITDDEGNAQIMQI 960  
 DB 901 ARTPEMTVTVLADVNDNPFVEFODEDFVEENSTIGLAVARVTAITDDEGNAQIMQI 960  
 QY 961 VEGNIPEVQOLDIFSGELTALVDLYEDRPEVYVIAQTSAPLVSRAVTHVRLDRNDP 1020  
 DB 961 VEGNIPEVQOLDIFSGELTALVDLYEDRPEVYVIAQTSAPLVSRAVTHVRLDRNDP 1020  
 QY 1021 PVLGHEFLFNNTYTNRSSPPGCAIGRVPAHDPIISLSTYSPFRGHELSVLINASTG 1080  
 DB 1021 PVLGHEFLFNNTYTNRSSPPGCAIGRVPAHDPIISLSTYSPFRGHELSVLINASTG 1080  
 QY 1081 ELKLSRALDNNRPLEAIVSVDGHSVTAQCALRVITIDEMLTHTSITLRLEDMSPER 1140  
 DB 1081 ELKLSRALDNNRPLEAIVSVDGHSVTAQCALRVITIDEMLTHTSITLRLEDMSPER 1140  
 QY 1141 FLSPLLGLFIOVAATLTAATPDHVVVFNVDTPDAPGHHILNVSLSVGQPPGGGPPFL 1200  
 DB 1141 FLSPLLGLFIOVAATLTAATPDHVVVFNVDTPDAPGHHILNVSLSVGQPPGGGPPFL 1200  
 QY 1201 PSEBQOEELYNRSLTLTAISORVLPDDNITCLREPCENYRCVSLRPSDAPPIASS 1260  
 DB 1201 PSEBQOEELYNRSLTLTAISORVLPDDNITCLREPCENYRCVSLRPSDAPPIASS 1260  
 QY 1261 VLFPRIFHVGLRGCPGFTGDYCTEVDLCYSRCPGPHRCRREGGYTCLCDGVTG 1320  
 DB 1261 VLFPRIFHVGLRGCPGFTGDYCTEVDLCYSRCPGPHRCRREGGYTCLCDGVTG 1320  
 QY 1321 EHCEVASRGRCTPGVCNNGGTCVNLVVGFKCPCSDPEKPYQVTTSRFPAPHSITP 1380  
 DB 1321 EHCEVASRGRCTPGVCNNGGTCVNLVVGFKCPCSDPEKPYQVTTSRFPAPHSITP 1380  
 QY 1381 RGLAQRHFHTLALSTATERBGLLYNRRFNKHDPALEVTIOBVOULTSAGSTTVS 1440  
 DB 1381 RGLAQRHFHTLALSTATERBGLLYNRRFNKHDPALEVTIOBVOULTSAGSTTVS 1440  
 QY 1441 PFVPEGVSDGOMHTYQLKYNKPLIGOTGLPQSPSEOKAVAVTVVDCCTGTALRGSVLTG 1500  
 DB 1441 PFVPEGVSDGOMHTYQLKYNKPLIGOTGLPQSPSEOKAVAVTVVDCCTGTALRGSVLTG 1500  
 QY 1501 NYSCAAQGTGGSKSLDLTGELLGVPDLPESPFVMMROFVGCMLNAYDSRIIDAD 1560  
 DB 1501 NYSCAAQGTGGSKSLDLTGELLGVPDLPESPFVMMROFVGCMLNAYDSRIIDAD 1560  
 QY 1561 FIANNGTVPCCPAKKNVCDNSTCHNGGTCVNOMDAFSCCEPLGRGSKCAQBMANPOHFL 1620  
 DB 1561 FIANNGTVPCCPAKKNVCDNSTCHNGGTCVNOMDAFSCCEPLGRGSKCAQBMANPOHFL 1620  
 QY 1621 GSSLVAMHGLSLPTISQPMYLSIMFRTROADGVLLOAITRGRSTITLQIRBGMVLSVGT 1680  
 DB 1621 GSSLVAMHGLSLPTISQPMYLSIMFRTROADGVLLOAITRGRSTITLQIRBGMVLSVGT 1680  
 QY 1681 GLQASSLALEPGRANDGDMHAQALGASGGPBAHILSFYQGOQABENLPRHLGHLIS 1740  
 DB 1681 GLQASSLALEPGRANDGDMHAQALGASGGPBAHILSFYQGOQABENLPRHLGHLIS 1740  
 QY 1741 NITVGGIIPGAPGAVARGGLQGVRSVDTPEGVNSLDPHGESINVOGGLDPPCCSN 1800  
 DB 1741 NITVGGIIPGAPGAVARGGLQGVRSVDTPEGVNSLDPHGESINVOGGLDPPCCSN 1800  
 QY 1801 PCPANSYCSNMDVSQSCDPRYGDNCTNVCDLNPFCEHOSVCTKPKSAPHGTYCBPCPN 1860  
 DB 1801 PCPANSYCSNMDVSQSCDPRYGDNCTNVCDLNPFCEHOSVCTKPKSAPHGTYCBPCPN 1860  
 QY 1861 YLGPYCETRIDQPCRGWGHPTGCPNCNCDVSKGFPDCKNTSGECHKENHYPGSPPT 1920  
 DB 1861 YLGPYCETRIDQPCRGWGHPTGCPNCNCDVSKGFPDCKNTSGECHKENHYPGSPPT 1920  
 QY 1921 CLDCDCYPTGSLSRVCEBDQCPCKPQVIGRQCDRCNPNPAEYVTTNGCEVNYDSCPAI 1980

DB 1921 CLDCDCYPTGSLSRVCEBDQCPCKPQVIGRQCDRCNPNPAEYVTTNGCEVNYDSCPAI 1980  
 QY 1981 EAGIWMPTRFGLPAAAPCPKSGFVVRHCDERHGMPLPULFNCTSTTFSELGFAERL 2040  
 DB 1981 EAGIWMPTRFGLPAAAPCPKSGFVVRHCDERHGMPLPULFNCTSTTFSELGFAERL 2040  
 QY 2041 QRNESGLDSRSGOQALILRNATQTAGVFGSDVKAVALTRLLAHSTORGFLSATQ 2100  
 DB 2041 QRNESGLDSRSGOQALILRNATQTAGVFGSDVKAVALTRLLAHSTORGFLSATQ 2100  
 QY 2101 DVHFTENILRVGSLTLTANKRHELIQOTREGTAMLLQHYEAYASALAOQMRHTYLSF 2160  
 DB 2101 DVHFTENILRVGSLTLTANKRHELIQOTREGTAMLLQHYEAYASALAOQMRHTYLSF 2160  
 QY 2161 TIVTENVIVSVRRJDKNFGAKLPRYEALRGBOPLDETIVILPESVFRTPPVVRPAG 2220  
 DB 2161 TIVTENVIVSVRRJDKNFGAKLPRYEALRGBOPLDETIVILPESVFRTPPVVRPAG 2220  
 QY 2221 PGEAQBEBELARRQRHPELSQGEAVASVITRTIAGLLPHNYDPDKRSLRVPKRPINT 2280  
 DB 2221 PGEAQBEBELARRQRHPELSQGEAVASVITRTIAGLLPHNYDPDKRSLRVPKRPINT 2280  
 QY 2281 PVVSIIVHDBELLPRALDKPVTVQFRLLEBETKPICVFMNHSILVSGTGMASAGCE 2340  
 DB 2281 PVVSIIVHDBELLPRALDKPVTVQFRLLEBETKPICVFMNHSILVSGTGMASAGCE 2340  
 QY 2341 VVFNESHVSQCNHMTSPAVLMDVSRRENGEILPLTLTVVAGVTLAALLTFPELTL 2400  
 DB 2341 VVFNESHVSQCNHMTSPAVLMDVSRRENGEILPLTLTVVAGVTLAALLTFPELTL 2400  
 QY 2401 LRLIRSNQHGIRNTTAAAGLAQVFLGGINQADLPACTVIAILLHPLYCTESMALLE 2460  
 DB 2401 LRLIRSNQHGIRNTTAAAGLAQVFLGGINQADLPACTVIAILLHPLYCTESMALLE 2460  
 QY 2461 ALHLRYALTEVRDNTGPMRFTYMLGNGVPAFTTGLAVGDDPEYGNPDCKMLSITYTLI 2520  
 DB 2461 ALHLRYALTEVRDNTGPMRFTYMLGNGVPAFTTGLAVGDDPEYGNPDCKMLSITYTLI 2520  
 QY 2521 WSFAGPVAFAVMSVFLYITLAARASCAORGFEEKKPVSGLOPSFVLLILSTWLLAL 2580  
 DB 2521 WSFAGPVAFAVMSVFLYITLAARASCAORGFEEKKPVSGLOPSFVLLILSTWLLAL 2580  
 QY 2581 LSVNSDTLLFHYLPATNCIOGPEIFLSYVLSKVRKALKACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHYLPATNCIOGPEIFLSYVLSKVRKALKACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSVNCSPYADGRLYQPYGDSAGSLHSTRSGKSPSYIPFLIRRESALNPGGPPGLG 2700  
 DB 2641 TSSVNCSPYADGRLYQPYGDSAGSLHSTRSGKSPSYIPFLIRRESALNPGGPPGLG 2700  
 QY 2701 DPGSLFLEGDDQCHDPTDSDLSLEDDQSGVASTHSDSEBEEBEEBAAFPDEOG 2760  
 DB 2701 DPGSLFLEGDDQCHDPTDSDLSLEDDQSGVASTHSDSEBEEBEEBAAFPDEOG 2760  
 QY 2761 WDSLIGFAERLPLHSTPKDGPGRGKAPWGPDTTAKSSGNGADBEERLRENGDALSR 2820  
 DB 2761 WDSLIGFAERLPLHSTPKDGPGRGKAPWGPDTTAKSSGNGADBEERLRENGDALSR 2820  
 QY 2821 EGSIGPLPGSSAOPHKIILKKKCLPTISSEKSLILRLPLEOCTGSSRRSSASBEGRGPP 2880  
 DB 2821 EGSIGPLPGSSAOPHKIILKKKCLPTISSEKSLILRLPLEOCTGSSRRSSASBEGRGPP 2880  
 QY 2881 RPPRQSLQOLNGVMPIMASIKAGTVDEBSSGSEFLPFNFILH 2923  
 DB 2881 RPPRQSLQOLNGVMPIMASIKAGTVDEBSSGSEFLPFNFILH 2923

RESULT 5  
 us-10-292-798-932  
 ; Sequence 932, Application us/10292798  
 ; Publication No. US2003023583A1  
 ; GENERAL INFORMATION:

APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 932  
LENGTH: 2923  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-292-798-932

Query Match 100.0%; Score 2923; DB 12; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLLDPOVGPGRSLSGRSGSACAPMGMLCPSS 60  
DB 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLLDPOVGPGRSLSGRSGSACAPMGMLCPSS 60

QY 61 SASNLMLYTSRCGADAGTELTGHLVPHHDLRWCPSESAHILPAPAPGCGPMSCLLGIG 120  
DB 61 SASNLMLYTSRCGADAGTELTGHLVPHHDLRWCPSESAHILPAPAPGCGPMSCLLGIG 120

QY 121 GHLSPOGKLTLPBEPCLAPRLCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180  
DB 121 GHLSPOGKLTLPBEPCLAPRLCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180

QY 181 POPPSQATVPENOPAGTVAISRAIDPDEGEAGLEVTMDALFDSRSNOFSLDPVGA 240  
DB 181 POPPSQATVPENOPAGTVAISRAIDPDEGEAGLEVTMDALFDSRSNOFSLDPVGA 240

QY 241 VTTAEELDEBETKSTHYFRVTAODHGMRRSALATLTLLVTDNHDPVFEQGEYKESLRE 300  
DB 241 VTTAEELDEBETKSTHYFRVTAODHGMRRSALATLTLLVTDNHDPVFEQGEYKESLRE 300

QY 301 NLEFVGYEVLTVRATGDAPPNANILYRLIEGSGSPSEVFEIDPRSGVIRTRGPDVREEV 360  
DB 301 NLEFVGYEVLTVRATGDAPPNANILYRLIEGSGSPSEVFEIDPRSGVIRTRGPDVREEV 360

QY 361 ESYVLTVASDQGRDGPSTTAAVLSVEEDNDNAPORSEKRYVQVREEDVTPGAPVLR 420  
DB 361 ESYVLTVASDQGRDGPSTTAAVLSVEEDNDNAPORSEKRYVQVREEDVTPGAPVLR 420

QY 421 VTASDRKSNVAVHYSIMSGNARQFYLDAGTALDVSPLDYETTKETTLRVARQDGG 480  
DB 421 VTASDRKSNVAVHYSIMSGNARQFYLDAGTALDVSPLDYETTKETTLRVARQDGG 480

QY 481 RPPLSVSGELVTVQVLDINDNAPIFVSTPPQATVLESVPLGYLVTHVQALDADAGNARL 540  
DB 481 RPPLSVSGELVTVQVLDINDNAPIFVSTPPQATVLESVPLGYLVTHVQALDADAGNARL 540

QY 541 EYRLAGVGHDPPTINGGWTGVSAAELDBREZDFPSPGVZEARDHOTPALTSASVSTV 600  
DB 541 EYRLAGVGHDPPTINGGWTGVSAAELDBREZDFPSPGVZEARDHOTPALTSASVSTV 600

QY 601 LDVNDNPPFTPOPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYOITSQNTNRSPITS 660  
DB 601 LDVNDNPPFTPOPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYOITSQNTNRSPITS 660

QY 661 QSGGGVSLALPLDYTLERQYVLAATVASTGTRODTQIYVNTDANTHRPVFQSSHYTN 720  
DB 661 QSGGGVSLALPLDYTLERQYVLAATVASTGTRODTQIYVNTDANTHRPVFQSSHYTN 720

QY 721 VNEDRAGTTVVLSATDEDTGENARITVFMEDESIQPIRIDADTGAVTTOQAEILDVEDQVS 780

DB 721 VNEDRAGTTVVLSATDEDTGENARITVFMEDESIQPIRIDADTGAVTTOQAEILDVEDQVS 780

QY 781 YTLAATDANGIPKSDTTYEILVNDVNDNAPQGLRDSYOGSYVEDVPFVSQISAT 840  
DB 781 YTLAATDANGIPKSDTTYEILVNDVNDNAPQGLRDSYOGSYVEDVPFVSQISAT 840

QY 841 DRDGLNGRVETTPQGGDGDGDFIVESTSGIVTLRLDRENNVAQYVLRVAVDKNMP 900  
DB 841 DRDGLNGRVETTPQGGDGDGDFIVESTSGIVTLRLDRENNVAQYVLRVAVDKNMP 900

QY 901 ARTPEVTVVLDVNDNPPVFEQDBPVFVBENSPICGLAVARVTRTDEDEGNQIMQI 960  
DB 901 ARTPEVTVVLDVNDNPPVFEQDBPVFVBENSPICGLAVARVTRTDEDEGNQIMQI 960

QY 961 VEGNIPFVFOIDIPSGELTALVDIDYDRPEYVLYOATSAPLVSRATVHYRLDRNDP 1020  
DB 961 VEGNIPFVFOIDIPSGELTALVDIDYDRPEYVLYOATSAPLVSRATVHYRLDRNDP 1020

QY 1021 PVLGNFEILFNNTYTNRSSFPFGAIGRPAPHDPIDSDLTYSFPERGNELSLVLLNASTG 1080  
DB 1021 PVLGNFEILFNNTYTNRSSFPFGAIGRPAPHDPIDSDLTYSFPERGNELSLVLLNASTG 1080

QY 1081 ELKLSRALDNNRPLEATMSVTVSDGHSVTACALRTVITIDEMLTHTSLRLBMSRER 1140  
DB 1081 ELKLSRALDNNRPLEATMSVTVSDGHSVTACALRTVITIDEMLTHTSLRLBMSRER 1140

QY 1141 FLSPLLGLFIQAVAAATLATPPDHVVVFNQDRTPAGGHIINVSISVQPPGPGGPPFL 1200  
DB 1141 FLSPLLGLFIQAVAAATLATPPDHVVVFNQDRTPAGGHIINVSISVQPPGPGGPPFL 1200

QY 1201 PSEDIQERLYINRSLTLAISQAVLPFDNTCLREPCENYRCVSLRPDSSAPFIASS 1260  
DB 1201 PSEDIQERLYINRSLTLAISQAVLPFDNTCLREPCENYRCVSLRPDSSAPFIASS 1260

QY 1261 VLFRPIHVGGLRCRCPGFTGDYCEFEVDLCYSRPGCPHRCRSGEGYCLCDGVTG 1320  
DB 1261 VLFRPIHVGGLRCRCPGFTGDYCEFEVDLCYSRPGCPHRCRSGEGYCLCDGVTG 1320

QY 1321 EHCEVSARSGRCTPGVCNKGTCVNLVGGFKCDPCPSGDFEKPYCOVTRSFPAHSFTTF 1380  
DB 1321 EHCEVSARSGRCTPGVCNKGTCVNLVGGFKCDPCPSGDFEKPYCOVTRSFPAHSFTTF 1380

QY 1381 RGLRQRFHTLALSPATFERDGLLYNGRFNEKDFVALBYIYOBQVLTFSAGESTTVS 1440  
DB 1381 RGLRQRFHTLALSPATFERDGLLYNGRFNEKDFVALBYIYOBQVLTFSAGESTTVS 1440

QY 1441 PVPBGVSDGQMHVYQLKYRKPLIGQTLPOGSEQKVAVTVDGCDTGVALARGSYLG 1500  
DB 1441 PVPBGVSDGQMHVYQLKYRKPLIGQTLPOGSEQKVAVTVDGCDTGVALARGSYLG 1500

QY 1501 NVSCAAGTQGGSKKSLDITGPLLGGVPLPESFPVAMROFVGCNMTQVDSRHIDNAD 1560  
DB 1501 NVSCAAGTQGGSKKSLDITGPLLGGVPLPESFPVAMROFVGCNMTQVDSRHIDNAD 1560

QY 1561 FIANNGTVPCCPAKKNVCDSTNGNGTCVNQMDAFSCBCLGFGGKSQAQEMANPOHFL 1620  
DB 1561 FIANNGTVPCCPAKKNVCDSTNGNGTCVNQMDAFSCBCLGFGGKSQAQEMANPOHFL 1620

QY 1621 GSSLVAMHGLSLPSQPMYLSLMFRTRQADGVLLQALTRGSTTTLQRLBGMVMSVGT 1680  
DB 1621 GSSLVAMHGLSLPSQPMYLSLMFRTRQADGVLLQALTRGSTTTLQRLBGMVMSVGT 1680

QY 1681 GLQASSTLRLBGRANDGDMHAQULAGASGGPBGHALLSPDYGOQRAEGNLPRLHGLHS 1740  
DB 1681 GLQASSTLRLBGRANDGDMHAQULAGASGGPBGHALLSPDYGOQRAEGNLPRLHGLHS 1740

QY 1741 NITVGGIIGPAGVARGRGCLQGYRVSDTPEGVNSLDPBSGESINVOGCSLPPCCSN 1800  
DB 1741 NITVGGIIGPAGVARGRGCLQGYRVSDTPEGVNSLDPBSGESINVOGCSLPPCCSN 1800

QY 1801 PCPANSYCSNDMSVSCSDGPGYGDNCTNVCNLPCEHGVCTRKPAAPGTYCCECPN 1860

```

Db 1801 PCPANSYCNMDYSVSCDPRYGDNCNVCNLPCEHOSVCTRKPAPDHGTCECPN 1860
QY 1861 YLGPYCETRI1DOFCBPGMWHPTGPONCDVSKGPPDCKNTSGECHKENHRRPPSP 1920
Db 1861 YLGPYCETRI1DOFCBPGMWHPTGPONCDVSKGPPDCKNTSGECHKENHRRPPSP 1920
QY 1921 CILCDYPTGSLSRVCDPEDGQPCPKGYIGROCDKCNPPAVTNGCEVNTDSCPRAI 1980
Db 1921 CILCDYPTGSLSRVCDPEDGQPCPKGYIGROCDKCNPPAVTNGCEVNTDSCPRAI 1980
QY 1981 BAGIWMPTRTFGLPALAAPCPKSGFGTAVRHCDHRGMLPNNLPNCTSIIPSELKGAERL 2040
Db 1981 BAGIWMPTRTFGLPALAAPCPKSGFGTAVRHCDHRGMLPNNLPNCTSIIPSELKGAERL 2040
QY 2041 QNENSGDGSQQLALLRNATQHTAGYFGSDVKAAYOLATRLAHSTORGFGLSAQ 2100
Db 2041 QNENSGDGSQQLALLRNATQHTAGYFGSDVKAAYOLATRLAHSTORGFGLSAQ 2100
QY 2101 DVHFTENILRVGSALLDTANKHWEILQOTEGTANLQHTAAYASALANMRHTYLSPF 2160
Db 2101 DVHFTENILRVGSALLDTANKHWEILQOTEGTANLQHTAAYASALANMRHTYLSPF 2160
QY 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGOPDLFTTYTILPESVRETPPVVRPAG 2220
Db 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGOPDLFTTYTILPESVRETPPVVRPAG 2220
QY 2221 PGEAOPBEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLARYKPIINT 2280
Db 2221 PGEAOPBEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLARYKPIINT 2280
QY 2281 PIVSISVHDEBELPRALDKPVTVOFRLLETERTPICVFNHSLIVSGGMSARGCE 2340
Db 2281 PIVSISVHDEBELPRALDKPVTVOFRLLETERTPICVFNHSLIVSGGMSARGCE 2340
QY 2341 VVFRSHSVSCQNMHTSPAVLMDVSRKNGELLPKLTLYVALGTLAALLTTPFELTL 2400
Db 2341 VVFRSHSVSCQNMHTSPAVLMDVSRKNGELLPKLTLYVALGTLAALLTTPFELTL 2400
QY 2401 IRLISNQHIGIRRNITAAALGALQVFLGINOADLPFACVTIATILHFLYCTFSWALIE 2460
Db 2401 IRLISNQHIGIRRNITAAALGALQVFLGINOADLPFACVTIATILHFLYCTFSWALIE 2460
QY 2461 ALHLYRALTEVDVNTGMPRTYMLGMPAPITGLAVLDPBGKGNPFCWLSTYDTLI 2520
Db 2461 ALHLYRALTEVDVNTGMPRTYMLGMPAPITGLAVLDPBGKGNPFCWLSTYDTLI 2520
QY 2521 WSPAPVAVAVMSVFLYTLAARASCAAROGFBEKKGPVSGLOPSPAVLILSATMLAL 2580
Db 2521 WSPAPVAVAVMSVFLYTLAARASCAAROGFBEKKGPVSGLOPSPAVLILSATMLAL 2580
QY 2581 LSVNSDTLLFHYLPAICNCIQQPPIFLSYVVLSEKVRKALKACSRKSPDPALTTKSTL 2640
Db 2581 LSVNSDTLLFHYLPAICNCIQQPPIFLSYVVLSEKVRKALKACSRKSPDPALTTKSTL 2640
QY 2641 TSSNYCPSRYAAGRLYOPYGDASGSLHSTRSGKQPSYIIFLLABESALNCGQPPGIG 2700
Db 2641 TSSNYCPSRYAAGRLYOPYGDASGSLHSTRSGKQPSYIIFLLABESALNCGQPPGIG 2700
QY 2701 DPGSLFLBEGDOQHDPDSDSLSLBDOQSGSYASTHSDSEBEEBEEBAAAPGSGG 2760
Db 2701 DPGSLFLBEGDOQHDPDSDSLSLBDOQSGSYASTHSDSEBEEBEEBAAAPGSGG 2760
QY 2761 WDSLLGPGAERLPLHSTPDKGPGPGKAPWPDFTTAKSSGNGAPBEERLRENDALSR 2820
Db 2761 WDSLLGPGAERLPLHSTPDKGPGPGKAPWPDFTTAKSSGNGAPBEERLRENDALSR 2820
QY 2821 BGSILGPLGSSAOPBKGLKKKCLPTISKSLALLPLEQCTGSSRGSSASGSGGPP 2880
Db 2821 BGSILGPLGSSAOPBKGLKKKCLPTISKSLALLPLEQCTGSSRGSSASGSGGPP 2880
QY 2881 RPPPROSLOEQNLGVPILAMSIKAGTVDEDSGSEFLFNFPLH 2923
Db 2881 RPPPROSLOEQNLGVPILAMSIKAGTVDEDSGSEFLFNFPLH 2923

```

```

RESULT 6
US-10-225-567A-524
; Sequence 524, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glenn C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 524
; LENGTH: 2923
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-524

```

```

Query Match 100.0%; Score 2923; DB 15; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSRPATGYPLPTPPPLLLLLLLLPPLLDQVGPCKSLGSGSSGACAPMGLCP 60
Db 1 MSRPATGYPLPTPPPLLLLLLLLPPLLDQVGPCKSLGSGSSGACAPMGLCP 60
QY 61 SASNMLYTSRCRDAGTELTGHIVPHDGLRVWCESRAHILPLPAPGCPMGCBLIG 120
Db 61 SASNMLYTSRCRDAGTELTGHIVPHDGLRVWCESRAHILPLPAPGCPMGCBLIG 120
QY 121 GHLSPOGRLTPEBHPCLKAPRLRCQSCKLAQAPGLARAGERSPEESLGRRKRNNTAPQ 180
Db 121 GHLSPOGRLTPEBHPCLKAPRLRCQSCKLAQAPGLARAGERSPEESLGRRKRNNTAPQ 180
QY 181 FOPPSYQATVBENOPAGTPVASLRAIDPDEBAGRLTYMDALPDSRNOFSLDPVTGA 240
Db 181 FOPPSYQATVBENOPAGTPVASLRAIDPDEBAGRLTYMDALPDSRNOFSLDPVTGA 240
QY 241 VTTABELRETKSTHVPFVTAQDHGMPRRSALATLTILVTQNDHPVFBQOEYKESLRE 300
Db 241 VTTABELRETKSTHVPFVTAQDHGMPRRSALATLTILVTQNDHPVFBQOEYKESLRE 300
QY 301 NLEVGVEVLTVRATDGDAPPNANILYRLBGSQGSPESEFVIDPRSGVITRTGPDVREEV 360
Db 301 NLEVGVEVLTVRATDGDAPPNANILYRLBGSQGSPESEFVIDPRSGVITRTGPDVREEV 360
QY 361 ESYQLTVEASQOGDPGGRSTTAAVFLSVEDNDNAPQPSSEKRYVQVREDEVTPGAPVLR 420
Db 361 ESYQLTVEASQOGDPGGRSTTAAVFLSVEDNDNAPQPSSEKRYVQVREDEVTPGAPVLR 420
QY 421 VTASDRDGSNAVHYHISMSGNARQGYFLDAQGALDVVSLDVEETTEYTLRPAADGG 480
Db 421 VTASDRDGSNAVHYHISMSGNARQGYFLDAQGALDVVSLDVEETTEYTLRPAADGG 480
QY 481 RPPLSNVGLTVTVQVLDINDNAPIFVSTPFOATVLESPVGLVLAHQVQADADAGDNARL 540
Db 481 RPPLSNVGLTVTVQVLDINDNAPIFVSTPFOATVLESPVGLVLAHQVQADADAGDNARL 540
QY 541 EYRLAGVGHDPPTTINNGTGWISVAABLDREEDVPFISGVARBHGTPALTAASVSTV 600
Db 541 EYRLAGVGHDPPTTINNGTGWISVAABLDREEDVPFISGVARBHGTPALTAASVSTV 600
QY 601 LDVVDNNPTFQPEYTVRLNEDAAVGTSVVVSAVDRASHVITYTQISGTRRFRSITS 660
Db 601 LDVVDNNPTFQPEYTVRLNEDAAVGTSVVVSAVDRASHVITYTQISGTRRFRSITS 660

```

QY 661 QSGGGVSLALPDLBYGLERQYVLAATASGTRDTAQIVVNTDANTHRPVFQSSHTTN 720  
 DB 661 QSGGGVSLALPDLBYGLERQYVLAATASGTRDTAQIVVNTDANTHRPVFQSSHTTN 720  
 QY 721 VNEBRPAGTIVVLAISDEDTGENARITVFMEDSIQFRIIDADTGAVTTOAELDYEDQVS 780  
 DB 721 VNEBRPAGTIVVLAISDEDTGENARITVFMEDSIQFRIIDADTGAVTTOAELDYEDQVS 780  
 QY 781 YTLATATARDNGIPQKSDITYLEILVNDVANDNAPOFLRDSYQGSVEDEVPFTSVLQISAT 840  
 DB 781 YTLATATARDNGIPQKSDITYLEILVNDVANDNAPOFLRDSYQGSVEDEVPFTSVLQISAT 840  
 QY 841 DEDSGANGVFTTFOGDDGDDGDFIYESTSGIVRTLRDLRENVAYVLAAYDXKMP 900  
 DB 841 DEDSGANGVFTTFOGDDGDDGDFIYESTSGIVRTLRDLRENVAYVLAAYDXKMP 900  
 QY 901 AATPEMVTATVLDVNDNPPVPEODEPFDVVEENSPIGLAVARTATPDGNAQIMYOI 960  
 DB 901 AATPEMVTATVLDVNDNPPVPEODEPFDVVEENSPIGLAVARTATPDGNAQIMYOI 960  
 QY 961 VEGNIPEVFOLDIFSGELTALVDLDYEDRPEYVLAQATSAPLVSRATVHRLDRNDP 1020  
 DB 961 VEGNIPEVFOLDIFSGELTALVDLDYEDRPEYVLAQATSAPLVSRATVHRLDRNDP 1020  
 QY 1021 PVLGNFEILFNNTYTRSSSPFGGALGRVPAHDPOISDITTSFERGNEILSVLNASTG 1080  
 DB 1021 PVLGNFEILFNNTYTRSSSPFGGALGRVPAHDPOISDITTSFERGNEILSVLNASTG 1080  
 QY 1081 ELKLSALNNRPLBAIMSVLSDGVHSTAOALVTIITDEMLHSTILRLBDSMPER 1140  
 DB 1081 ELKLSALNNRPLBAIMSVLSDGVHSTAOALVTIITDEMLHSTILRLBDSMPER 1140  
 QY 1141 FLSPLILGFIQAAVATLATPPDHVVFNVQRTDABGHIILNVSLVGQPPGSGGPPPL 1200  
 DB 1141 FLSPLILGFIQAAVATLATPPDHVVFNVQRTDABGHIILNVSLVGQPPGSGGPPPL 1200  
 QY 1201 PEBDQBERLYNRSLTLTASQAVLPFDNICLRECEYVNCVSLARDSSAPFLASS 1260  
 DB 1201 PEBDQBERLYNRSLTLTASQAVLPFDNICLRECEYVNCVSLARDSSAPFLASS 1260  
 QY 1261 VLFPRPIHPVGLRCRCPFGTGDYCTEYDLCYSRCGPHGRCSRSGGTCICRDGYG 1320  
 DB 1261 VLFPRPIHPVGLRCRCPFGTGDYCTEYDLCYSRCGPHGRCSRSGGTCICRDGYG 1320  
 QY 1321 EHCEVASRSGRCTPGVCKNGGTCVNLVGGFKCDPSGDPBEKPYCQVTRSPFASFTF 1380  
 DB 1321 EHCEVASRSGRCTPGVCKNGGTCVNLVGGFKCDPSGDPBEKPYCQVTRSPFASFTF 1380  
 QY 1381 RGLRORFHFTLALSPATKRDGLLYNGBPNKHDVALVIOBOVQLTFSGESTTYS 1440  
 DB 1381 RGLRORFHFTLALSPATKRDGLLYNGBPNKHDVALVIOBOVQLTFSGESTTYS 1440  
 QY 1441 PVPBGVSOGMHTVQKTKYKPLLGQTGLPOGSPBQKAVAVTVDGCDGVALRFGSVLG 1500  
 DB 1441 PVPBGVSOGMHTVQKTKYKPLLGQTGLPOGSPBQKAVAVTVDGCDGVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQGSKSLDTGLPLLLGGVDPBPSPFVNRQFVGNCRNLQVDSRHITMAD 1560  
 DB 1501 NYSCAAQGTQGSKSLDTGLPLLLGGVDPBPSPFVNRQFVGNCRNLQVDSRHITMAD 1560  
 QY 1561 FLANNNTVGCPRKKNVCSNCHNGCTCVNOMDASCECPILGFGKSCAQEBANQHL 1620  
 DB 1561 FLANNNTVGCPRKKNVCSNCHNGCTCVNOMDASCECPILGFGKSCAQEBANQHL 1620  
 QY 1621 GSSLVAMHGLSLPISOPWYLSLMTFRTOAGVLLQAITRGRSTITTLQLEBHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISOPWYLSLMTFRTOAGVLLQAITRGRSTITTLQLEBHVMSVEGT 1680  
 QY 1681 GLQASSTLRLEPRANDGDWHALALGASGGPHAILSPYGOORABGNLGPRLHGLHLS 1740  
 DB 1681 GLQASSTLRLEPRANDGDWHALALGASGGPHAILSPYGOORABGNLGPRLHGLHLS 1740  
 QY 1741 NITVGGIPBAGVAGFRGCTLGVRVSDTPBEGVNSLDPBHGESINVEQCSLPDPCSN 1800

DB 1741 NITVGGIPBAGVAGFRGCTLGVRVSDTPBEGVNSLDPBHGESINVEQCSLPDPCSN 1800  
 QY 1801 PCPANSYCSNMDVSQSCDPOXYGDNCTNYCDLNPCEHOSQCTRKPSAPHGTCCECPN 1860  
 DB 1801 PCPANSYCSNMDVSQSCDPOXYGDNCTNYCDLNPCEHOSQCTRKPSAPHGTCCECPN 1860  
 QY 1861 YLGPYCETRIDQPCBGMWGHPTGPCNCDVSKGPDPCNKTSGECHKENHTRPGSPPT 1920  
 DB 1861 YLGPYCETRIDQPCBGMWGHPTGPCNCDVSKGPDPCNKTSGECHKENHTRPGSPPT 1920  
 QY 1921 CLICDCTYGSLSRVCDEDDQCPCKREVIGRQCCRCNNPABVYTNCEVNYVDSCPAL 1980  
 DB 1921 CLICDCTYGSLSRVCDEDDQCPCKREVIGRQCCRCNNPABVYTNCEVNYVDSCPAL 1980  
 QY 1981 BAGIWMPTRRFGLPAAACPKPSFGTARHCDERHGMPLPMLFNCTSIPTFSELKGFARL 2040  
 DB 1981 BAGIWMPTRRFGLPAAACPKPSFGTARHCDERHGMPLPMLFNCTSIPTFSELKGFARL 2040  
 QY 2041 QNBSGLDSGSSQOLALLRNATOHTAGYFGSDVAVAYQALATRLAHSTORGFLSATQ 2100  
 DB 2041 QNBSGLDSGSSQOLALLRNATOHTAGYFGSDVAVAYQALATRLAHSTORGFLSATQ 2100  
 QY 2101 DVHFTENILRVGSALLDPANRHWELIOOTEGGTAMLOHYEAYSAALANMRHTYLSPT 2160  
 DB 2101 DVHFTENILRVGSALLDPANRHWELIOOTEGGTAMLOHYEAYSAALANMRHTYLSPT 2160  
 QY 2161 TIVTPNIVISVVRLDKGNFAGAKLPRYBALRGQPPDLETTVILPESVFRFETPPVVRAG 2220  
 DB 2161 TIVTPNIVISVVRLDKGNFAGAKLPRYBALRGQPPDLETTVILPESVFRFETPPVVRAG 2220  
 QY 2221 PGEAOPEBELARORRHPLSQGEAVASVIIYRTLAGLLPHNYDDKXSLRVKPKPIINT 2280  
 DB 2221 PGEAOPEBELARORRHPLSQGEAVASVIIYRTLAGLLPHNYDDKXSLRVKPKPIINT 2280  
 QY 2281 PUVSISYHDDDELPRALDKPTVOFRLJTBERTKPCVPMNHSILVSGGMSARGCE 2340  
 DB 2281 PUVSISYHDDDELPRALDKPTVOFRLJTBERTKPCVPMNHSILVSGGMSARGCE 2340  
 QY 2341 VVFRNESHVSCCNHMTSPAVLMVDSRRNGEILPLKLTJYVALGVTLLAALLTFEFLT 2400  
 DB 2341 VVFRNESHVSCCNHMTSPAVLMVDSRRNGEILPLKLTJYVALGVTLLAALLTFEFLT 2400  
 QY 2401 LRIILSNQHIRNUTALGLAQLVFLGINOADIPTACTYAILHFLYCTFEWALLE 2460  
 DB 2401 LRIILSNQHIRNUTALGLAQLVFLGINOADIPTACTYAILHFLYCTFEWALLE 2460  
 QY 2461 ALHLRYALTBRYDVNTGPMRFYMLGMVPAFITGLAVGLDPBEGYNDPCWLSIYDTLI 2520  
 DB 2461 ALHLRYALTBRYDVNTGPMRFYMLGMVPAFITGLAVGLDPBEGYNDPCWLSIYDTLI 2520  
 QY 2521 WSPAGPVAFAVMSVFLYTLAARASCAAOQOFKKGVSGLQPSFAVILLLSATWLLAL 2580  
 DB 2521 WSPAGPVAFAVMSVFLYTLAARASCAAOQOFKKGVSGLQPSFAVILLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLFTKNCIGPPTLSYVYLSKXVRKALKLACGRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLFTKNCIGPPTLSYVYLSKXVRKALKLACGRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNCSPYADGRLYOPYGSAGSLHSTSRGKSQPSYIPTLAEBSALNPGQPPGLG 2700  
 DB 2641 TSSYNCSPYADGRLYOPYGSAGSLHSTSRGKSQPSYIPTLAEBSALNPGQPPGLG 2700  
 QY 2701 DPGSLFLBGQDQHDPDJDSDSLLEDDQGSYVASTHSDBSEEBEEREAAPGEG 2760  
 DB 2701 DPGSLFLBGQDQHDPDJDSDSLLEDDQGSYVASTHSDBSEEBEEREAAPGEG 2760  
 QY 2761 WSSLGPAABERLPLHSTKQGGPGPKAPWPCDPCCTTKKSSGNAPEERLRENDALSR 2820  
 DB 2761 WSSLGPAABERLPLHSTKQGGPGPKAPWPCDPCCTTKKSSGNAPEERLRENDALSR 2820  
 QY 2821 ESSLGPLGSSAQPKGILLKKCLPTISEKSLJLPLQCTGSSGSSASGSGGPP 2880

Db	2821	EGSLGPLESSAQPRIKGLIKKCLPTISKSLSLPLLEQCTGSSRGSSASESGRGPP	2880
Qy	2881	RPPRQSLQEQNGVMPIMASIKAGTVDESSGSEFLPFENFLH	2923
Db	2881	RPPRQSLQEQNGVMPIMASIKAGTVDESSGSEFLPFENFLH	2923
RESULT 7			
US-09-788-711A-2			
Sequence 2, Application US/09788711A			
Patent No. US20020058328A1			
GENERAL INFORMATION:			
APPLICANT: Tania Tamein Teata			
TITLE OF INVENTION: NOVEL COMPOUNDS			
FILE REFERENCE: GP-30225			
CURRENT APPLICATION NUMBER: US/09/788,711A			
CURRENT FILING DATE: 2001-02-20			
PRIOR APPLICATION NUMBER: 0004196.2			
PRIOR FILING DATE: 2000-02-19			
NUMBER OF SEQ ID NOS: 4			
SOFTWARE: FASTSEQ for Windows Version 3.0			
SEQ ID NO 2			
LENGTH: 2956			
TYPE: PRT			
ORGANISM: HOMO SAPIENS			
US-09-788-711A-2			
Query Match			
Beat Local Similarity 97.1%; Score 2837; DB 9; Length 2956;			
Matches 2837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MRSPATGVPPLPPPPPLLLLLLLLPPLLDQVPCPSLSGRGSSGACAPMGLCP	60
Db	1	MRSPATGVPPLPPPPPLLLLLLLLPPLLDQVPCPSLSGRGSSGACAPMGLCP	60
Qy	61	SASINMLVTSRCRDAGTELTGHLVPHHGLRWCESEAHIPLPAPPGCCPMSCLLIG	120
Db	61	SASINMLVTSRCRDAGTELTGHLVPHHGLRWCESEAHIPLPAPPGCCPMSCLLIG	120
Qy	121	GHLSPQGLTLPEEHPCLKAPRLRCQSCLAQAPGLRAGERSPEBSLGGRRKANYTAPO	180
Db	121	GHLSPQGLTLPEEHPCLKAPRLRCQSCLAQAPGLRAGERSPEBSLGGRRKANYTAPO	180
Qy	181	FOPEYQATVPENQAGTPVASLRAIDPDEGAGLEETMDALPDSRNQPFSLDPYGA	240
Db	181	FOPEYQATVPENQAGTPVASLRAIDPDEGAGLEETMDALPDSRNQPFSLDPYGA	240
Qy	241	VTTAEELREKTSHTVFVTAQDHGMPRRSALA TLITVTDPNDHPVPEQETKESLRE	300
Db	241	VTTAEELREKTSHTVFVTAQDHGMPRRSALA TLITVTDPNDHPVPEQETKESLRE	300
Qy	301	NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFELIDPSGVI RTGSPVDRSEV	360
Db	301	NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFELIDPSGVI RTGSPVDRSEV	360
Qy	361	ESYQLTVEASQGRDPGRSTTAAVFLSVEDNDNAPQPSKRYVYVQREDPGAPVLR	420
Db	361	ESYQLTVEASQGRDPGRSTTAAVFLSVEDNDNAPQPSKRYVYVQREDPGAPVLR	420
Qy	421	VTAASRDGSAVVYHYSIMSGNARGOFTLDAQTGALDVVSPLDVETTKETLRYAAQDGG	480
Db	421	VTAASRDGSAVVYHYSIMSGNARGOFTLDAQTGALDVVSPLDVETTKETLRYAAQDGG	480
Qy	481	RPPLSNVGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGLVLAQAIADAGDNARL	540
Db	481	RPPLSNVGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGLVLAQAIADAGDNARL	540
Qy	541	EYRLAGVHADPPTTINNGTGMISVAEELDREVDVYSGVEARDHGTPALTAASVSATV	600
Db	541	EYRLAGVHADPPTTINNGTGMISVAEELDREVDVYSGVEARDHGTPALTAASVSATV	600
Qy	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRDASHVITYQTISGNTRRRPFSTS	660
Db	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRDASHVITYQTISGNTRRRPFSTS	660
Db	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRDASHVITYQTISGNTRRRPFSTS	660
Qy	661	OSGGELVSLALPLDYKLEROVLAVTASDGTROPTAOI VVWVTANTRPVPOSSHVYN	720
Db	661	OSGGELVSLALPLDYKLEROVLAVTASDGTROPTAOI VVWVTANTRPVPOSSHVYN	720
Qy	721	VNEBRPACTTVLISATDEDTGENARITYEMEDSIPOFRIDADGAVTTQAELEDVQVS	780
Db	721	VNEBRPACTTVLISATDEDTGENARITYEMEDSIPOFRIDADGAVTTQAELEDVQVS	780
Qy	781	YTLAITANDNGIPKSDTTYELIVVNDVNDAPQFLRDSYGSYVEDVPPFTSVLOISAT	840
Db	781	YTLAITANDNGIPKSDTTYELIVVNDVNDAPQFLRDSYGSYVEDVPPFTSVLOISAT	840
Qy	841	DRDGLANGRVYFTQGGDDGDPFVSESTGIVTLRLRDEENVAQYVLRAYADKGNP	900
Db	841	DRDGLANGRVYFTQGGDDGDPFVSESTGIVTLRLRDEENVAQYVLRAYADKGNP	900
Qy	901	ARTPMEVTVTLVDVNDNPVPEODEPVDVFEENSPIGLAVARVATDPDEGTNAQIMYQI	960
Db	901	ARTPMEVTVTLVDVNDNPVPEODEPVDVFEENSPIGLAVARVATDPDEGTNAQIMYQI	960
Qy	961	VEGINIPERVOLDISGELTALVDLDYEDRPETVYVQTASAPLYSRATVHYRLDRNDP	1020
Db	961	VEGINIPERVOLDISGELTALVDLDYEDRPETVYVQTASAPLYSRATVHYRLDRNDP	1020
Qy	1021	PVLGNFELFNNTYNNSSSPFGAIGRVPAHDDISDLTYSFERNELSLVYLNASTG	1080
Db	1021	PVLGNFELFNNTYNNSSSPFGAIGRVPAHDDISDLTYSFERNELSLVYLNASTG	1080
Qy	1081	ELKLSRALDNNRPLEAIVSVSDVHSAVTQAQALRYTITIDEMUTHSITLRLDMSPER	1140
Db	1081	ELKLSRALDNNRPLEAIVSVSDVHSAVTQAQALRYTITIDEMUTHSITLRLDMSPER	1140
Qy	1141	FLSPLLGLFTQAVAAATLATPPDHVNVNVCQDTAPGCHILNVSLSGQPPGCGGPFEL	1200
Db	1141	FLSPLLGLFTQAVAAATLATPPDHVNVNVCQDTAPGCHILNVSLSGQPPGCGGPFEL	1200
Qy	1201	PSEDLQERLYNRLSLTAISAQVLPDPDNI CLAREPCENYRCVSVLRFDSAPFIASS	1260
Db	1201	PSEDLQERLYNRLSLTAISAQVLPDPDNI CLAREPCENYRCVSVLRFDSAPFIASS	1260
Qy	1261	VLFRPIHPVGGALRCRCPGFTGDYCETEVDLCYSRPGCPHRCRSREGGYTCLCRDGYTG	1320
Db	1261	VLFRPIHPVGGALRCRCPGFTGDYCETEVDLCYSRPGCPHRCRSREGGYTCLCRDGYTG	1320
Qy	1321	EHCEVSARSRCRTGVCNKGCTCNVLLVSGFKCIBSGSDPEKPYCQVTTTRSFPASHFTTF	1380
Db	1321	EHCEVSARSRCRTGVCNKGCTCNVLLVSGFKCIBSGSDPEKPYCQVTTTRSFPASHFTTF	1380
Qy	1381	RGLRQRHFHTLASFATKERDGLLYNGRFRNEKHDPALEVIQBOVOLTFSSAGESTTVS	1440
Db	1381	RGLRQRHFHTLASFATKERDGLLYNGRFRNEKHDPALEVIQBOVOLTFSSAGESTTVS	1440
Qy	1441	PFEVGGVSDQWHTVOLKTYNKPILLGQTGLPQGSSEQKVAVVTVDGCDDTVALRFGSVLG	1500
Db	1441	PFEVGGVSDQWHTVOLKTYNKPILLGQTGLPQGSSEQKVAVVTVDGCDDTVALRFGSVLG	1500
Qy	1501	NYSQAQGTQGGSKSLDITGPIILGVPDLPSFPVRMQPVCMMNLQVDSRHIMAD	1560
Db	1501	NYSQAQGTQGGSKSLDITGPIILGVPDLPSFPVRMQPVCMMNLQVDSRHIMAD	1560
Qy	1561	FIANGTVPGCPAKONVCDSENTCHNGGT CYNOMDAFCECEPLGFGGSKCAQEMANPOHFL	1620
Db	1561	FIANGTVPGCPAKONVCDSENTCHNGGT CYNOMDAFCECEPLGFGGSKCAQEMANPOHFL	1620
Qy	1621	GSSIVANHGSLPISQPMWYLSLMEFTRQADGVLLQAITRRGRSTITTLQREGHVLSEGT	1680
Db	1621	GSSIVANHGSLPISQPMWYLSLMEFTRQADGVLLQAITRRGRSTITTLQREGHVLSEGT	1680
Qy	1681	GLQASSLRLRPGRANODMHAOLALGASGPGHAILISPYGOORAGCNLGPRLHGLHS	1740
Db	1681	GLQASSLRLRPGRANODMHAOLALGASGPGHAILISPYGOORAGCNLGPRLHGLHS	1740



```

QY 1741 NITVGGIPGPAGVAGRGFGCLQGVVSVTDPGVNSLDPGSHSINVBOGCSLPPDCSN 1800
DB 1741 NITVGGIPGPAGVAGRGFGCLQGVVSVTDPGVNSLDPGSHSINVBOGCSLPPDCSN 1800
QY 1801 PCPANSYCSNDMDYSVSCDPCPYGDNCCNVCDLNPCHQSVCTKPSAPHYTCBCPN 1860
DB 1801 PCPANSYCSNDMDYSVSCDPCPYGDNCCNVCDLNPCHQSVCTKPSAPHYTCBCPN 1860
QY 1861 YLGPYCETRIIDQPCPRGWMGHPTGFCNCQDVSKGFPDPCNKTSGECHCKENYRPPGSP 1920
DB 1861 YLGPYCETRIIDQPCPRGWMGHPTGFCNCQDVSKGFPDPCNKTSGECHCKENYRPPGSP 1920
QY 1921 CLLCDCYFPGSLSRVCDPBDGQCPCKPGVITGQCDCNDNPAEVTNNGCEVYDSCPRAI 1980
DB 1921 CLLCDCYFPGSLSRVCDPBDGQCPCKPGVITGQCDCNDNPAEVTNNGCEVYDSCPRAI 1980
QY 1981 BAGIWMPPRRFGJPAAPCPKGSFGFAVHCEHGMPLPNNLFCNCTSTPSLKKPAREL 2040
DB 1981 BAGIWMPPRRFGJPAAPCPKGSFGFAVHCEHGMPLPNNLFCNCTSTPSLKKPAREL 2040
QY 2041 QNBSGLDGRSQOQLALLRNATQHTAGYFGSDVAVAYQATRLLAHSTQGFGLSATQ 2100
DB 2041 QNBSGLDGRSQOQLALLRNATQHTAGYFGSDVAVAYQATRLLAHSTQGFGLSATQ 2100
QY 2101 DVHFTENLARSALIDTANKHWEILQOTEGSTAMLIQHYEAVASALAKNNRHTYLSPF 2160
DB 2101 DVHFTENLARSALIDTANKHWEILQOTEGSTAMLIQHYEAVASALAKNNRHTYLSPF 2160
QY 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEOPDLETTVILPESVRETRPPVVRPG 2220
DB 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEOPDLETTVILPESVRETRPPVVRPG 2220
QY 2221 PGEAGEPEELARORRHPELQGEAVASYIYRTLAGLPHNYDDPKSLARVKEPIINT 2280
DB 2221 PGEAGEPEELARORRHPELQGEAVASYIYRTLAGLPHNYDDPKSLARVKEPIINT 2280
QY 2281 PAVVISVHDDDELPRALDKPVTVQRLLETERKPCICVFNHSHLVSGTGMARQGE 2340
DB 2281 PAVVISVHDDDELPRALDKPVTVQRLLETERKPCICVFNHSHLVSGTGMARQGE 2340
QY 2341 VVFRSHSVSCCNHMTSPAVLMDVSRNGEELPLKLTLYVALGVTLAALLTFFPLTL 2400
DB 2341 VVFRSHSVSCCNHMTSPAVLMDVSRNGEELPLKLTLYVALGVTLAALLTFFPLTL 2400
QY 2401 IRLISNOHGIRRNLTAAALGAQLVFLGINOADLPFACTVIAIILHFLYLCFPMALL 2460
DB 2401 IRLISNOHGIRRNLTAAALGAQLVFLGINOADLPFACTVIAIILHFLYLCFPMALL 2460
QY 2461 AMLYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEGVGNPFCMLSTYDILI 2520
DB 2461 AMLYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEGVGNPFCMLSTYDILI 2520
QY 2521 WSPFAGVAVAVMSVFLYTLARASCAAOQGFEEKGVPVSGLOSPAVLILLSATWLAL 2580
DB 2521 WSPFAGVAVAVMSVFLYTLARASCAAOQGFEEKGVPVSGLOSPAVLILLSATWLAL 2580
QY 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLASYVLISKEVRKALKIACSRKSPDPAALTTKSTL 2640
DB 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLASYVLISKEVRKALKIACSRKSPDPAALTTKSTL 2640
QY 2641 TSSNCPSPYAGGRLYOPYGDASGSLHSTRSGKQSPSTIPFLARESLANGQGPGLG 2700
DB 2641 TSSNCPSPYAGGRLYOPYGDASGSLHSTRSGKQSPSTIPFLARESLANGQGPGLG 2700
QY 2701 DPGSLFLBEGODQHPDSTDLSLEDQSGSYASTHSDSEEBEEREAAPFGEG 2760
DB 2701 DPGSLFLBEGODQHPDSTDLSLEDQSGSYASTHSDSEEBEEREAAPFGEG 2760
QY 2761 WSLILGPAGERLPLHSTPDQGGPGKAPWPGDFTTAKSSGNGAPBERLRENDALSR 2820
DB 2761 WSLILGPAGERLPLHSTPDQGGPGKAPWPGDFTTAKSSGNGAPBERLRENDALSR 2820

```

```

QY 2821 EGSIGPLPGSSAQPHKG 2837
DB 2821 EGSIGPLPGSSAQPHKG 2837

RESULT 8
US-09-843-856-2
; Sequence 2, Application US/09843856
; Patent No. US20020034785A1
; GENERAL INFORMATION:
; APPLICANT: SOPEP, DANIEL R.
; LI, YI
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CALCITONIN RECEPTOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/843,856
; FILING DATE: 30-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/970,758
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STREPE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0660001/EKS/KMT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-843-856-2

Query Match 18.3%; Score 534; DB 9; Length 568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2388 LAALLTFFPLTLIRLSNOHGIRRNLTAAALGAQLVFLGINOADLPFACTVIAIILH 2447
DB 33 LAALLTFFPLTLIRLSNOHGIRRNLTAAALGAQLVFLGINOADLPFACTVIAIILH 92

QY 2448 FLVLCFSWMLLBALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEGVGN 2507
DB 93 FLVLCFSWMLLBALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEGVGN 152

QY 2508 PDFCMLSYDITLINSFAGPVAVAVMSVFLYTLARASCAAOQGFEEKGVPVSGLOSPFA 2567
DB 153 PDFCMLSYDITLINSFAGPVAVAVMSVFLYTLARASCAAOQGFEEKGVPVSGLOSPFA 212

QY 2568 VLLLSATWMLALSVNSDTLLFHYLPATCNCIOGPFIFLASYVLISKEVRKALKIACSRK 2627
DB 213 VLLLSATWMLALSVNSDTLLFHYLPATCNCIOGPFIFLASYVLISKEVRKALKIACSRK 272

QY 2628 PSPDPALTTKSTLTSYNCPSPYAGGRLYOPYGDASGSLHSTRSGKQSPSTIPFLAREE 2687
DB 273 PSPDPALTTKSTLTSYNCPSPYAGGRLYOPYGDASGSLHSTRSGKQSPSTIPFLAREE 332

```

QY 2688 SALNPGQPPGIGDPSGLFLEGDDQHDPTDSDSLIEDDQSGSYASTHSSDSEEEE 2747  
DB 333 SALNPGQPPGIGDPSGLFLEGDDQHDPTDSDSLIEDDQSGSYASTHSSDSEEEE 392  
QY 2748 EEEBAAPGEGQWMSILGPGARLPLHSTPKDGGPGKAPWPDFTTAKESSGNAP 2807  
DB 393 EEEBAAPGEGQWMSILGPGARLPLHSTPKDGGPGKAPWPDFTTAKESSGNAP 452  
QY 2808 EERLRNGDALSRSGSLGFLPGSSAQPHKGIKKKCLPTISEKSSILRLPLEQCTGSSRG 2867  
DB 453 EERLRNGDALSRSGSLGFLPGSSAQPHKGIKKKCLPTISEKSSILRLPLEQCTGSSRG 512  
QY 2868 SASSESGSRGPPPPPPROSLDQOLNGVPMIAMSITAGTVDEDSGSEFLPFNF 2921  
DB 513 SASSESGSRGPPPPPPROSLDQOLNGVPMIAMSITAGTVDEDSGSEFLPFNF 566

RESULT 9  
US-10-176-847-100  
; Sequence 100, Application US/10176847  
; Publication No. US20030068636A1  
; GENERAL INFORMATION:  
; APPLICANT: Veibry, Petter Ole  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST  
; TITLE OF INVENTION: AND OVARIAN CANCER  
; FILE REFERENCE: MRI-039  
; CURRENT APPLICATION NUMBER: US/10/176,847  
; CURRENT FILING DATE: 2002-06-21  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: PatSeq for Windows Version 4.0  
; SEQ ID NO 100  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-176-847-100

Query Match 16.6%; Score 485; DB 15; Length 565;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2430 INQADLPACTVIALHLHPLVCTSMALLLHLYRALTBRDVNTGPMRYMLGCV 2489  
DB 74 INQADLPACTVIALHLHPLVCTSMALLLHLYRALTBRDVNTGPMRYMLGCV 133  
QY 2490 PAFITGLAVGLDPEGYNDPFCMLSIYDTLINSFAGPVAFVMSVFLYIIAARASCAQ 2549  
DB 134 PAFITGLAVGLDPEGYNDPFCMLSIYDTLINSFAGPVAFVMSVFLYIIAARASCAQ 193  
QY 2550 ROGFEKGPVSGLDQSPAVLLIILSATWLLIALLSVNSDTLLPFLYLPATCNCIOGPFIFLSY 2609  
DB 194 ROGFEKGPVSGLDQSPAVLLIILSATWLLIALLSVNSDTLLPFLYLPATCNCIOGPFIFLSY 253  
QY 2610 VLSKRVKALKLACSRKSPDPALTTKSTLSSVNCSPYADGELYQPYGDSAGSLHST 2669  
DB 254 VLSKRVKALKLACSRKSPDPALTTKSTLSSVNCSPYADGELYQPYGDSAGSLHST 313  
QY 2670 SRSQSGQSYIFPLLRRESALNPGQPPGLPGSLFLEGDDQHDPTDSDSLIEDD 2729  
DB 314 SRSQSGQSYIFPLLRRESALNPGQPPGLPGSLFLEGDDQHDPTDSDSLIEDD 373  
QY 2730 QSGSYASTHSSDSEEEBAAPGEGWMSILGPGARLPLHSTPKDGGPGKAP 2789  
DB 374 QSGSYASTHSSDSEEEBAAPGEGWMSILGPGARLPLHSTPKDGGPGKAP 433  
QY 2790 WPGDGTAKESSGNAPGAPERLRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISE 2849  
DB 434 WPGDGTAKESSGNAPGAPERLRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISE 493  
QY 2850 KSSLRLPLEQCTGSSRGSSASGSGRGPAPPPOSLDQOLNGVPMIAMSITAGTVDE 2909  
DB 494 KSSLRLPLEQCTGSSRGSSASGSGRGPAPPPOSLDQOLNGVPMIAMSITAGTVDE 553

QY 2910 DSSGS 2914  
DB 554 DSSGS 558

RESULT 10  
US-10-264-237-2041  
; Sequence 2041, Application US/10264237  
; Publication No. US20040009491A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P431P1  
; CURRENT APPLICATION NUMBER: US/10/264,237  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/16450  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 60/205,515  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 2876  
; SOFTWARE: Patent In Ver. 3.1  
; SEQ ID NO 2041  
; LENGTH: 568  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (83)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
; NAME/KEY: MISC FEATURE  
; LOCATION: (240)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (522)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
US-10-264-237-2041

Query Match 9.6%; Score 281; DB 12; Length 568;  
Best Local Similarity 100.0%; Pred. No. 2.9e-210; Indels 0; Gaps 0;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2596 TCNCIOGPFIFLSYVLSKEVRKALKLACSRKSPDPALTTKSTLSSVNCSPYADGRL 2655  
DB 241 TCNCIOGPFIFLSYVLSKEVRKALKLACSRKSPDPALTTKSTLSSVNCSPYADGRL 300  
QY 2656 YQPYGDSAGSLHSTSRGSKQPSYIFPLLRRESALNPGQPPGLGDPGLFLEGDDQHD 2715  
DB 301 YQPYGDSAGSLHSTSRGSKQPSYIFPLLRRESALNPGQPPGLGDPGLFLEGDDQHD 360  
QY 2716 PPTSDSDSLIEDDQSGSYASTHSSDSEEEBAAPGEGWMSILGPGARLPLH 2775  
DB 361 PPTSDSDSLIEDDQSGSYASTHSSDSEEEBAAPGEGWMSILGPGARLPLH 420  
QY 2776 STPDGGPGKAPWPDFTTAKESSGNAPGAPERLRNGDALSRSGSLGFLPGSSAQPH 2835  
DB 421 STPDGGPGKAPWPDFTTAKESSGNAPGAPERLRNGDALSRSGSLGFLPGSSAQPH 480  
QY 2836 KGIKKKCLPTISEKSSILRLPLEQCTGSSRGSSASGSRG 2876  
DB 481 KGIKKKCLPTISEKSSILRLPLEQCTGSSRGSSASGSRG 521

RESULT 11  
US-09-925-300-1299  
; Sequence 1299, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1299  
LENGTH: 717  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (39)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (147)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (181)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (232)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (379)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (389)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (671)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1299

Query Match  
Best Local Similarity 100.0%; Score 281; DB 10; Length 717;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2596 TCNCIGPFIPLSYVLSKVRKALKACSRKSPDPALTTKSTLTSYNCSPFYADGR 2655  
DB 390 TCNCIGPFIPLSYVLSKVRKALKACSRKSPDPALTTKSTLTSYNCSPFYADGR 449

QY 2656 YQPYGDSAGSLHSTSGSKSQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOHD 2715  
DB 450 YQPYGDSAGSLHSTSGSKSQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOHD 509

QY 2716 PPTDSDSLSEDDQSGSYSTSSDSEEEEBEERAAFPQGGWDSLGPAGRLPLH 2775  
DB 510 PPTDSDSLSEDDQSGSYSTSSDSEEEEBEERAAFPQGGWDSLGPAGRLPLH 569

QY 2776 STPKDQGPGRKAPWQDPTGTAKSSGNGAPERRLRENDALSRGSLGPLGSSAOPH 2835  
DB 570 STPKDQGPGRKAPWQDPTGTAKSSGNGAPERRLRENDALSRGSLGPLGSSAOPH 629

QY 2836 KGIKKKCLPTISEKSSLLRLPLEQCTGSSRGSSAEGSRG 2876  
DB 630 KGIKKKCLPTISEKSSLLRLPLEQCTGSSRGSSAEGSRG 670

RESULT 12  
US-10-017-161-1096  
Sequence 1096, Application US/10017161  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 084335/0152

FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/10/017,161  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: JP 2001/246789  
PRIOR FILING DATE: 2001-06-16  
NUMBER OF SEQ ID NOS: 2430  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1096  
LENGTH: 646  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-017-161-1096

Query Match  
Best Local Similarity 100.0%; Score 179; DB 12; Length 646;  
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2601 QGPFIFLSYVLSKVRKALKACSRKSPDPALTTKSTLTSYNCSPFYADGR 2660  
DB 221 QGPFIFLSYVLSKVRKALKACSRKSPDPALTTKSTLTSYNCSPFYADGR 280

QY 2661 DSAGSLHSTSGSKSQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOHD 2720  
DB 281 DSAGSLHSTSGSKSQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOHD 340

QY 2721 DSDLSLEDDQSGSYSTSSDSEEEEBEERAAFPQGGWDSLGPAGRLPLH 2779  
DB 341 DSDLSLEDDQSGSYSTSSDSEEEEBEERAAFPQGGWDSLGPAGRLPLH 399

RESULT 13  
US-09-764-870-331  
Sequence 331, Application US/09764870  
Patent No. US20020042386A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT214  
CURRENT APPLICATION NUMBER: US/09/764,870  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 646  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 331  
LENGTH: 219  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (119)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (123)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (145)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (189)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (192)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (196)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (211)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (214)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (219)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-870-331

Query Match 4.0%; Score 118; DB 9; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.5e-83;  
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 60  
QY 627 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGLVSLALPLDYKLERQYVLA 684  
DB 61 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGLVSLALPLDYKLERQYVLA 118

## RESULT 14

US-10-125-540-331  
Sequence 331, Application US/10125540  
Publication No. US20030059875A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT214CI

CURRENT APPLICATION NUMBER: US/10/125,540

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 646

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 331

LENGTH: 219

TYPE: PR1

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (119)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (123)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (145)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (189)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (192)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (196)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (211)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (214)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:

NAME/KEY: misc feature

LOCATION: (219)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-125-540-331

Query Match 4.0%; Score 118; DB 15; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.5e-83;

Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 60  
QY 627 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGLVSLALPLDYKLERQYVLA 684  
DB 61 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGLVSLALPLDYKLERQYVLA 118

## RESULT 15

US-09-764-870-479  
Sequence 479, Application US/09764870  
Patent No. US20020042386A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT214

CURRENT APPLICATION NUMBER: US/09/764,870

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 646

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 479

LENGTH: 111

TYPE: PR1

ORGANISM: Homo sapiens

US-09-764-870-479

Query Match 3.4%; Score 100; DB 9; Length 111;  
Best Local Similarity 100.0%; Pred. No. 9e-70;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLDVNDNNPTFQPEYTVRLNEDAVG 60  
QY 627 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGL 666  
DB 61 TSVVTVSAVDRDAHSVITYQITSGNTRNRFSTISQSGGL 100

Search completed: February 11, 2004, 16:16:15  
Job time : 73 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:05:06 ; Search time 30 Seconds

(without alignments)  
4122,488 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923

Sequence: 1 MRSFATGVPLTPPPPLLL.....AGTVDSSGSEFLPFNFH 2923

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	0.7	44	1	US-07-998-003A-16
2	21	0.7	44	1	US-08-453-274B-16
3	21	0.7	44	1	US-08-453-695A-16
4	21	0.7	44	1	US-08-268-161A-16
5	21	0.7	44	2	US-08-453-702A-16
6	21	0.7	44	2	US-09-099-639-16
7	21	0.7	44	5	PCT-US93-12588-16
8	21	0.7	44	5	PCT-US95-08071-16
9	18	0.6	884	2	US-08-465-976A-2
10	18	0.6	884	2	US-08-982-412-2
11	15	0.5	43	1	US-07-998-003A-83
12	15	0.5	43	1	US-08-453-274B-83
13	15	0.5	43	1	US-08-453-695A-83
14	15	0.5	43	1	US-08-268-161A-83
15	15	0.5	43	2	US-08-453-702A-83
16	15	0.5	43	2	US-09-099-639-83
17	15	0.5	43	3	PCT-US93-12588-83
18	15	0.5	43	3	PCT-US95-08071-83
19	13	0.4	311	5	US-08-318-837-9
20	13	0.4	566	4	US-09-491-522-7
21	13	0.4	871	3	US-09-245-041-19
22	13	0.4	1211	4	US-09-491-522-5
23	13	0.4	1350	3	US-09-245-041-17
24	12	0.4	969	2	US-08-548-159-1
25	12	0.4	986	2	US-08-548-159-3
26	12	0.4	1012	4	US-08-811-481-16
27	12	0.4	1310	3	US-08-989-299-10

28	11	0.4	26	4	US-09-336-536-59	Sequence 59, Appl
29	11	0.4	44	1	US-07-998-003A-79	Sequence 79, Appl
30	11	0.4	44	1	US-08-453-274B-79	Sequence 79, Appl
31	11	0.4	44	1	US-08-453-695A-79	Sequence 79, Appl
32	11	0.4	44	1	US-08-268-161A-79	Sequence 79, Appl
33	11	0.4	44	2	US-08-453-702A-79	Sequence 79, Appl
34	11	0.4	44	3	US-09-099-639-79	Sequence 79, Appl
35	11	0.4	44	5	PCT-US93-12588-79	Sequence 79, Appl
36	11	0.4	44	5	PCT-US95-08071-79	Sequence 79, Appl
37	11	0.4	213	4	US-09-336-536-58	Sequence 58, Appl
38	11	0.4	492	1	US-07-794-393-4	Sequence 3, Appl
39	11	0.4	492	1	US-08-001-711-4	Sequence 4, Appl
40	11	0.4	739	3	US-09-022-983-2	Sequence 2, Appl
41	11	0.4	740	3	US-09-022-983-5	Sequence 2, Appl
42	11	0.4	1001	4	US-08-884-569A-2	Sequence 2, Appl
43	11	0.4	1646	4	US-09-535-008-67	Sequence 67, Appl
44	11	0.4	1647	4	US-09-535-008-2	Sequence 2, Appl
45	11	0.4				

#### ALIGNMENTS

##### RESULT 1

US-07-998-003A-16  
Sequence 16, Application US/07998003A

Patent No. 5643781

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESS: Marshall, O'Toole, Gerstein, Murray, &

ADDRESS: Bicknell

STREET: 20 South Clark Street

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/998,003A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5643781and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 30903

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/346-5750

TELEFAX: 312/984-9740

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 44 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-998-003A-16

Query Match 0.7%; Score 21; DB 1; Length 44;

Matches 21; Conservativity 100.0%; Pred. No. 5.4e-11; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVVDINDNAP 503  
DB 24 PLSNVSGLVTVQVVDINDNAP 44

RESULT 2  
US-08-453-274B-16  
Sequence 16, Application US/08453274B  
Patent No. 5663300  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-274B-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred.No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
Db 24 PLSNVSGLVTVQVLDINDNAP 44

RESULT 3  
US-08-453-695A-16  
Sequence 16, Application US/08453695A  
Patent No. 5708143  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,695A

FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-695A-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred.No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
Db 24 PLSNVSGLVTVQVLDINDNAP 44

RESULT 4  
US-08-268-161A-16  
Sequence 16, Application US/08268161A  
Patent No. 5798224  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,161A  
FILING DATE: June 27, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Young J. Sun  
REGISTRATION NUMBER: P-41,337  
REFERENCE/DOCKET NUMBER: 27866/32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-268-161A-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred.No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 5

US-08-453-702A-16  
Sequence 16, Application US/08453702A  
Patent No. 5891706  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-702A-16

Query Match 0.7%; Score 21; DB 2; Length 44;

Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 6

US-09-099-639-16  
Sequence 16, Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161  
FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 6262237and  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-099-639-16

## Query Match

0.7%; Score 21; DB 3; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 7

PCT-US93-12588-16  
Sequence 16, Application PC/TUS9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-12588-16

Query Match 0.7%; Score 21; DB 5; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSVSGLVTVQVLDINDNP 503  
DB 24 PLSVSGLVTVQVLDINDNP 44

## RESULT 8

PCT-US95-08071-16  
Sequence 16, Application PC/TUS9508071  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08071  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE: 23 DEC 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-08071-16

Query Match 0.7%; Score 21; DB 5; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSVSGLVTVQVLDINDNP 503  
DB 24 PLSVSGLVTVQVLDINDNP 44

## RESULT 9

US-08-465-976A-2  
Sequence 2, Application US/08465976A  
Patent No. 5869632  
GENERAL INFORMATION:  
APPLICANT: SOPPET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: RUBEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN GILFILLAN, CECCHI  
ADDRESSEE: STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NJ  
COUNTRY: US  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,976A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY F  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 994-1700  
TELEFAX: (201) 994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 884 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-976A-2

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGLDPFGYGNPDCW 2512  
DB 220 GLAVGLDPFGYGNPDCW 237

## RESULT 10

US-08-982-412-2  
Sequence 2, Application US/08982412  
Patent No. 5958729  
GENERAL INFORMATION:  
APPLICANT: SOPPET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: RUBEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/982,412  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF181PCT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 884 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-982-412-2

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3,9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGLDPEGYGNPDPCW 2512  
Db 220 GLAVGLDPEGYGNPDPCW 237

RESULT 11  
US-07-998-003A-83  
Sequence 83, Application US/07998003A  
Patent No. 5643781  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 20 South Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/998,003A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5643781and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 33903  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/346-5750  
TELEFAX: 312/984-9740  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-998-003A-83

Query Match 0.5%; Score 15; DB 1; Length 43;

Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 791 GIPOKSDTYTLEIIV 805  
Db 21 GIPOKSDTYTLEIIV 35

RESULT 12  
US-08-453-274B-83  
Sequence 83, Application US/08453274B  
Patent No. 5663300  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-274B-83

Query Match 0.5%; Score 15; DB 1; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GIPOKSDTYTLEIIV 805  
Db 21 GIPOKSDTYTLEIIV 35

RESULT 13  
US-08-453-695A-83  
Sequence 83, Application US/08453695A  
Patent No. 5708143  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453, 695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-83

Query Match          0.5%; Score 15; DB 1; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GIPKSDTYLEILV 805
DB 21 GIPKSDTYLEILV 35

RESULT 14
US-08-268-161A-83
; Sequence 83, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-268-161A-83

Query Match          0.5%; Score 15; DB 1; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GIPKSDTYLEILV 805
DB 21 GIPKSDTYLEILV 35

RESULT 15
US-08-453-702A-83
; Sequence 83, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,702A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5891706and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32657
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-702A-83

Query Match          0.5%; Score 15; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GIPKSDTYLEILV 805
DB 21 GIPKSDTYLEILV 35
```

Search completed: February 11, 2004, 16:09:43  
Job time : 31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:03:46 ; Search time 39 Seconds

(without alignments)  
7207.719 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923

Sequence: 1 MRSPATGVPLTPPPPLLL.....AGTVDEDSGSGSEFLPFNPLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

#### SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	23	0.8	3034	2	seven-pass transme
2	18	0.6	1364	2	MEGF2 protein - hu
3	12	0.4	491	2	stromelysin 3 (RC
4	12	0.4	608	2	hypothetical prote
5	12	0.4	1015	2	transmembrane tyro
6	12	0.4	1015	2	phogrin precursor
7	12	0.4	1032	2	guanylate kinase-i
8	12	0.4	1309	1	peptidyl-dipeptida
9	11	0.4	75	2	hypothetical prote
10	11	0.4	106	2	proline/leucine-ri
11	11	0.4	134	2	secretin precursor
12	11	0.4	182	2	receptor-activit
13	11	0.4	189	2	receptor activit
14	11	0.4	213	2	neuromodulin - gol
15	11	0.4	316	2	hypothetical prote
16	11	0.4	393	2	transcription init
17	11	0.4	412	2	transcription init
18	11	0.4	445	2	NNE-4AG - human (f
19	11	0.4	492	2	hypothetical prote
20	11	0.4	527	2	stromelysin 3 (RC
21	11	0.4	602	2	80K protein H prec
22	11	0.4	617	2	prostaglandin G/H
23	11	0.4	617	2	cyclooxygenase 1 -
24	11	0.4	675	2	actin-binding prot
25	11	0.4	699	2	MYD protein inhib
26	11	0.4	736	2	histidine rich cal
27	11	0.4	740	2	Fas-binding protei
28	11	0.4	768	2	Fas-binding protei
29	11	0.4	777	2	protein kinase (EC

30	11	0.4	777	2	F54024	protein kinase (EC
31	11	0.4	779	2	B54024	protein kinase (EC
32	11	0.4	783	2	A55817	cyclin-dependent k
33	11	0.4	1001	2	S30385	G9a protein - huma
34	11	0.4	1022	2	I53078	homeotic gene regu
35	11	0.4	1078	2	T42712	myelin transcripti
36	11	0.4	1184	2	A55184	fibulin-2 precursor
37	11	0.4	1467	2	T48162	hypothetical prote
38	11	0.4	1613	2	S39059	protein BRG1 - hum
39	11	0.4	1647	2	S45252	SNP2beta protein -
40	11	0.4	1753	2	T00350	hypothetical prote
41	11	0.4	2009	2	S49764	SKC7 protein - yea
42	10	0.3	81	2	T48398	hypothetical prote
43	10	0.3	81	2	T47289	hypothetical prote
44	10	0.3	105	2	A27471	homeotic protein R
45	10	0.3	119	2	A03314	homeotic protein m

#### ALIGNMENTS

##### RESULT 1

T14119

seven-pass transmembrane receptor protein precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14119

R:Hadjantonakis, A.K.; Formstone, C.J.; Little, P.F.R.

Submitted to the EMBL Data Library, October 1997

A/Description: The Celst family of novel evolutionarily conserved seven-pass transmembran

A/Reference number: Z17881

A/Accession: T14119

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-3034 <HAD>

A/Cross-References: EMBL:AF031572; NID:G3800735; PID:G3800736; PIDN:AAC6836.1

C/Genetics:

A/Gene: Celst1

A/Map position: 15

C/Keywords: transmembrane protein

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-3034/Product: seven-pass transmembrane receptor protein #status predicted <MAY>

Query Match 0.8%; Score 23; DB 2; Length 3034;

Best Local Similarity 100.0%; Pred. No. 9.3e-12;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1219 ISARVLPFDNICTLRPCENYM 1241

Db 1309 ISARVLPFDNICTLRPCENYM 1331

##### RESULT 2

T00250

MEGF2 protein - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000

C/Accession: T00250

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomic 5', 27-34, 1998

A/Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs

A/Reference number: Z14126; MUID:98360089; PMID:9693030

A/Accession: T00250

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1364 <YAK>

A/Cross-References: EMBL:AB011536; NID:G3449297; PIDN:BA32464.1; PID:G3449298

A/Experimental source: brain; Clone HG1044

C/Genetics:

A/Gene: MEGF2

A/Map position: 3p21.2-p24.1

C/Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homo

F:1-28/Domain: EGF homology (fragment) <EGF>

F:32-66/Domain: EGF homology <EGF1>  
F:124-169/Domain: laminin-type EGF-like homology <LEG>

Query Match 0.4%; Score 18; DB 2; Length 1364;  
Best Local Similarity 100.0%; Pred. No. 2e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGLDPGGYGNPDFCM 2512  
DB 710 GLAVGLDPGGYGNPDFCM 727

RESULT 3  
JC6197  
stromelysin 3 (EC 3.4.24.-) - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 17-Mar-1999  
C/Accession: JC6197  
R:Okada, A.; Saez, S.; Mismul, Y.; Bassett, P.  
Gene 185, 187-193, 1997  
A>Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin an  
A:Reference number: JC6197; MUID:97208872; PMID:9055814  
A:Contents: Skin wounds  
A:Accession: JC6197  
A:Molecule type: mRNA  
A:Residues: 1-491 <OKA>  
A:Cross-references: GB:U46034  
C:Comment: This protein is a member of the matrix metalloproteinase family.  
C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloprotei  
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F:52-261/Domain: matrix metalloproteinase homology <MMP>  
F:291-483/Domain: hemopexin repeat homology <HPN>  
F:84,218,222,228/Binding site: zinc, catalytic (His, His, His) (inhibited) #status  
F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
F:219/Active site: Glu #status predicted

Query Match 0.4%; Score 12; DB 2; Length 491;  
Best Local Similarity 100.0%; Pred. No. 0.03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PILLILLILLLP 27  
DB 20 PILLILLILLLP 31

RESULT 4  
T06632  
hypothetical protein T20K18.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
C/Accession: T06632  
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkee, W.; Stiekema, W.; Bancroft, I.; Men  
submitted to the Protein Sequence Database, April 1999  
A:Reference number: Z15790  
A:Accession: T06632  
A:Molecule type: DNA  
A:Residues: 1-608 <BEV>  
A:Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.90  
C:Experimental source: cultivar Columbia; BAC clone T20K18  
C:Genetic:  
A:Gene: ATSP:T20K18.90  
A:Map position: 4  
A:Introns: 202/3; 216/2; 221/3; 241/3; 292/2; 471/1

Query Match 0.4%; Score 12; DB 2; Length 608;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2742 SEEBEEREEA 2753  
DB 97 SEEBEEREEA 108

RESULT 5  
JC5263  
transmembrane tyrosine phosphatase-like protein, ICAR - human

C:Species: Homo sapiens (man)  
C>Date: 25-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 20-Jun-2000  
C/Accession: JC5263  
R:Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.  
Biochem. Biophys. Res. Commun. 229, 402-411, 1996  
A>Title: ICAR, a novel member of a new family of transmembrane, tyrosine phosphatase-11  
A:Reference number: JC5263; MUID:97127415; PMID:8954911  
A:Accession: JC5263  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1015 <SMI>  
A:Cross-references: GB:Y08569; NID:G1644377; PID:CA69880.1; PID:G1644378  
C:Comment: This protein has an intracellular protein tyrosine phosphatase like protein.  
C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata  
C:Keywords: phosphoprotein  
F:770-994/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:945/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:951/Binding site: substrate phosphate (Arg) #status predicted

Query Match 0.4%; Score 12; DB 2; Length 1015;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PILLILLILLLP 27  
DB 6 PILLILLILLLP 17

RESULT 6  
JC5062  
phogrin precursor - human  
N:Contains: protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type  
C:Species: Homo sapiens (man)  
C>Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 21-Jun-2002  
C/Accession: JC5062; T46903  
R:Kawasaki, E.; Hutton, J.C.; Eisenbach, G.S.  
Biochem. Biophys. Res. Commun. 227, 440-447, 1996  
A>Title: Molecular cloning and characterization of the human transmembrane protein tyros  
A:Reference number: JC5062; MUID:97032784; PMID:8878534  
A:Contents: Islet  
A:Accession: JC5062  
A:Molecule type: mRNA  
A:Residues: 1-1015 <KAW>  
A:Cross-references: GB:U6707; NID:G1620663; PID:AA050742.1; PID:G1620664  
R:Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.  
Biochem. Biophys. Res. Commun. 229, 402-411, 1996  
A>Title: ICAR, a novel member of a new family of transmembrane, tyrosine phosphatase-11  
A:Reference number: JC5263; MUID:97127415; PMID:8954911  
A:Accession: JC5263  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-207, 'S', 209-246, 'G', 248-322, 'R', 324, 'N', 326-1015 <SMI>  
A:Cross-references: GB:Y08569; NID:G1644377; PID:CA69880.1; PID:G1644378  
R:Amoroge, W.; Wilkner, U.; Mewes, H.W.; Weill, B.; Wiemann, S.  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24134  
A:Accession: T46903  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 'DA', 714, 771-1015 <AAA>  
A:Cross-references: EMBL:AL157451  
C:Comment: This protein has an intracellular protein tyrosine phosphatase like protein.  
C:Genetic:  
A:Note: DKP2G761A0712.1  
C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-1015/Product: phogrin #status predicted <MAT>  
F:615-639/Domain: transmembrane #status predicted <TM>



F:770-994/Domain: protein-tyrosine-phosphatase homology <PTP2>  
F:945/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F:951/Binding site: substrate phosphate (Arg) #status predicted

Query Match 0.4%; Score 12; DB 2; Length 1015;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 PLLLLLLLLLP 27  
DB 6 PLLLLLLLLLP 17

RESULT 7  
T18293  
guanylate kinase-interacting protein 1 Maguin-1, membrane-associated - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18293  
R:Yao, L.; Hata, Y.; Takai, Y.  
Submitted to the EMBL Data Library, October 1998  
A:Description: Membrane-associated guanylate kinase-interacting protein.  
A:Reference number: Z18862  
A:Accession: T18293

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1032 <YAO>  
A:Cross-references: EMBL:AF102853; NID:G4151804; PID:G4151805; PIDN:AA04567.1

Query Match 0.4%; Score 12; DB 2; Length 1032;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2743 EEEEEEEEA 2754  
DB 877 EEEEEEEEA 888

RESULT 8  
S35484  
peptidyl-dipeptidase A (EC 3.4.15.1) precursor, pulmonary splice form - rabbit  
N:Alternate names: angiotensin-converting enzyme; dipeptidyl carboxypeptidase I; kinnase  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S35484; A23455; A18700; A38655; A49726; S17509  
R:Thekkumkara, T.J.; Livingston III, W.; Kumar, R.S.; Sen, G.C.  
Nucleic Acids Res. 20, 683-687, 1992  
A:Title: Use of alternative polyadenylation sites for tissue-specific transcription of t  
A:Reference number: S35484; MUID:92178960; PMID:1311831  
A:Accession: S35484  
A:Molecule type: mRNA  
A:Residues: 1-1309 <THR>  
A:Cross-references: EMBL:X62551  
R:Wata, K.; Blacher, R.; Soffer, R.L.; Lai, C.Y.  
Arch. Biochem. Biophys. 227, 188-201, 1983  
A:Reference number: A23455; MUID:84051289; PMID:6314908  
A:Accession: A23455  
A:Molecule type: protein  
A:Residues: 34-47, 'N', 49-55 <IWA>  
A:Experimental source: lung  
R:Wata, K.; Lai, C.Y.; El-Dorri, H.A.; Soffer, R.L.  
Biochem. Biophys. Res. Commun. 107, 1097-1103, 1982  
A:Title: The NH2- and COOH-terminal sequences of the angiotensin-converting enzyme isozy  
A:Reference number: A90107; MUID:83048249; PMID:6291514  
A:Accession: A18700

A:Molecule type: protein  
A:Residues: 34-44,754-755, 'U', 757 <IWA>  
R:Kumar, R.S.; Thekkumkara, T.J.; Sen, G.C.  
J. Biol. Chem. 266, 3854-3862, 1991  
A:Title: The mRNAs encoding the two angiotensin-converting isozyms are transcribed from  
A:Reference number: A38655; MUID:91139683; PMID:1847388  
A:Accession: A38655  
A:Molecule type: DNA

A:Residues: 1-88 <KDM>  
A:Cross-references: GB:M58579  
R:Ranchandran, R.; Sen, G.C.; Misano, K.; Sen, I.  
J. Biol. Chem. 269, 2125-2130, 1994  
A:Title: Regulated cleavage-secretion of the membrane-bound angiotensin-converting enzyme  
A:Reference number: A49726; MUID:94124566; PMID:8294466  
A:Accession: A49726

A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1236-1258 <RAM>  
A:Experimental source: testis  
R:Kirtley, T.L.  
Biochem. J. 278, 375-380, 1991  
A:Title: The Mg(2+)-ATPase of rabbit skeletal-muscle transverse tubule is a highly glycos  
A:Reference number: S17509; MUID:91378880; PMID:1654880  
A:Accession: S17509

A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 34-55 <KIR>  
A:Comment: This enzyme converts angiotensin I to angiotensin II in presence of divalent c  
ver, the enzyme has been found also in renal tubules and intestinal mucosa.  
C:Superfamily: mammalian peptidyl-dipeptidase A  
C:Keywords: alternative splicing; blood pressure control; chloride; glycoprotein; interest  
P:1-33/Domain: signal sequence #status predicted <SIG>  
P:34-1309/Product: peptidyl-dipeptidase A, pulmonary #status experimental <MAT>  
P:59,79,150,322,448,512,680,698,717,945,1194/Binding site: carbohydrate (Asn) (covalent)

Query Match 0.4%; Score 12; DB 1; Length 1309;  
Best Local Similarity 100.0%; Pred. No. 0.07;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 PPPLLLLLLL 24  
DB 15 PPPLLLLLLL 26

RESULT 9  
T06013  
hypothetical protein T25K17.100 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
C:Accession: T06013  
R:Bayan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15382  
A:Accession: T06013  
A:Molecule type: DNA  
A:Residues: 1-75 <BRV>  
A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.100  
A:Experimental source: cultivar Columbia; BAC clone T25K17  
C:Genetics:  
A:Gene: ATSP:T25K17.100  
A:Map position: 4

Query Match 0.4%; Score 11; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 0.052;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2743 EEEEEEEEA 2753  
DB 13 EEEEEEEEA 23

RESULT 10  
T06479  
proline/leucine-rich protein precursor - garden pea  
C:Species: Pisum sativum (garden pea)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Nov-1999  
C:Accession: T06479  
R:Rodriguez-Concepcion, M.; Perez-Garcia, A.; Beltran, J.  
submitted to the EMBL Data Library, November 1995  
A:Description: Isolation of cDNAs which accumulate during pea (Pisum sativum L.) early fr  
A:Reference number: Z15708

```

A:Accession: T06479
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-106 <R0D>
A:Cross-references: EMBL:267873; NID:g1213129; PIDN:CAA1780.1; PID:e208986
F1-1//Domain: signal sequence #status predicted <Sig>
F18-106/Product: proline/leucine-rich protein #status predicted <MAT>

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 106;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LLLLLPPPL 30
DB 73 LLLLLPPPL 83

RESULT 11
A:Accession: A40959
A:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 16-Jul-1999
C:Accession: A40886; A40959; A35094; A32544
R:Itton, N.; Furuya, T.; Ozaki, K.; Onita, W.; Kawasaki, T.
J. Biol. Chem. 266, 12595-12598, 1991
A:Title: The secretin precursor gene. Structure of the coding region and expression in t
A:Reference number: A40886; MUID:91286291; PMID:2061329
A:Accession: A40886
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-134 <RTO>
A:Cross-references: GB:M63984; NID:g206889; PIDN:AAA42127.1; PID:g206890
R:Kopin, A.S.; Wheeler, M.B.; Nishitani, J.; MCBride, E.W.; Chang, T.; Chey, W.Y.; Lett
Proc. Natl. Acad. Sci. U.S.A. 88, 5335-5339, 1991
A:Title: The secretin gene: evolutionary history, alternative splicing, and development
A:Reference number: A40959; MUID:91271364; PMID:1711228
A:Accession: A40959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-134 <R0P>
A:Cross-references: GB:M64033; NID:g206891; PIDN:AAA42128.1; PID:g206892
R:Kopin, A.S.; Wheeler, M.B.; Lettler, A.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2289-2303, 1990
A:Title: Secretin: structure of the precursor and tissue distribution of the mRNA.
A:Reference number: A35094; MUID:90192795; PMID:2315322
A:Accession: A35094
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-134 <R0P2>
A:Cross-references: GB:M31495; NID:g206887; PIDN:AAA42126.1; PID:g206888
R:Gossein, D.; Vandermereers, A.; Vandermereers-Piret, M.C.; Rathe, J.; Cauvin, A.; Robberech
Biochem. Biophys. Res. Commun. 160, 862-867, 1989
A:Title: Isolation and primary structure of rat secretin.
A:Reference number: A32544; MUID:89246545; PMID:2719704
A:Accession: A32544
A:Status: preliminary
A:Molecule type: protein
A:Residues: 33-59 <GOS>
C:Superfamily: glucagon
C:Keyword: duplication

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 134;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25
DB 8 PLLLLLLLLL 18

RESULT 12
JC7236
receptor-activity-modifying protein 2 - rat

```

```

C:Species: Rattus norvegicus (Norway rat)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: JC7236
R:Nagase, T.; Mkyama, M.; Sugawara, A.; Mori, K.; Yahata, K.; Kasahara, M.; Suganami, T
Biochem. Biophys. Res. Commun. 270, 89-93, 2000
A:Title: Rat receptor-activity-modifying proteins (RAMPs) for adrenomedullin/CGRP recept
A:Reference number: JC7235
A:Accession: JC7236
A:Molecule type: mRNA
A:Residues: 1-182 <NAG>
A:Cross-references: DDBJ:AB030943
C:Comment: This protein is a transmembrane glycoprotein having roles in renal pathophysic
as the adrenomedullin receptor.
C:Keywords: glycoprotein; transmembrane protein

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 182;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25
DB 29 PLLLLLLLLL 39

RESULT 13
JC7262
receptor activity modifying protein 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: JC7262
R:Ono, Y.; Okano, I.; Kojima, M.; Okada, K.; Kangawa, K.
Biochem. Biophys. Res. Commun. 271, 197-202, 2000
A:Title: Decreased gene expression of adrenomedullin receptor in mouse lungs during seps
A:Reference number: JC7261
A:Accession: JC7262
A:Molecule type: mRNA
A:Residues: 1-189 <ONO>
A:Cross-references: GB:AF209906
C:Comment: This protein interacts with calcitonin receptor-like receptor, modifies its lik
lation.
C:Genetics:
A:Gene: Yamp2
C:Keywords: transmembrane protein

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 189;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25
DB 29 PLLLLLLLLL 39

RESULT 14
UQ0075
neuromodulin - goldfish
N:Alternate names: B-50; F1; GAP-43; PP46
C:Species: Carassius auratus (goldfish)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jul-2000
C:Accession: UQ0075
R:Labate, M.E.; Skene, J.H.P.
Neuron 3, 299-310, 1989
A:Title: Selective conservation of GAP-43 structure in vertebrate evolution.
A:Reference number: UQ0075; MUID:90380372; PMID:2641999
A:Accession: UQ0075
A:Molecule type: mRNA
A:Residues: 1-213 <LAB>
A:Cross-references: GB:M26250; NID:g212953; PIDN:AAA03010.1; PID:g212954
A>Note: this protein contains potential sites for fatty acylation and membrane attachment
C:Comment: GAP-43 is a major growth cone membrane protein whose expression is widely cor
C:Comment: Features of the protein are most highly conserved in vertebrate evolution.
C:Comment: The amino end (1-57) has a strictly conserved domain including protein-protein
C:Superfamily: histone H1

```

C:Keywords: calmodulin binding; lipoprotein; phosphoprotein; thiolester bond  
 F:3,4/Binding site: palmitate (Cys) (covalent) #status predicted  
 F:81/Binding site: phosphate (Ser) (covalent) #status predicted

## Query Match

0.4%; Score 11; DB 2; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 EEEEEEEEA 2753  
 |||||

DB 161 EEEEEEEEA 171

## RESULT 15

G96513

hypothetical protein T3P24.8 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: G96513

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Malt, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96513

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:AB05173; NID:g9993350; PIDN:AG11423.1; GSPDB:GN00141

C:Genetics:

A:Gene: T3P24.8

A:Map position: 1

## Query Match

0.4%; Score 11; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2742 SEEEEEEE 2752  
 |||||

DB 220 SEEEEEEE 230

Search completed: February 11, 2004, 16:09:00  
 Job time : 40 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:58:30 ; Search time 25 Seconds

(without alignments)  
5498.362 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923  
Sequence: 1 MRSPTATGVLPTPPPLLL.....AGTVDEDSGSRFLPFNPLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	2923	1	CLR2_HUMAN
2	187	6.4	2144	1	CLR2_RAT
3	168	5.7	2920	1	CLR1_MOUSE
4	23	0.8	3034	1	CLR1_MOUSE
5	20	0.7	3014	1	CLR1_HUMAN
6	18	0.6	3301	1	CLR3_MOUSE
7	18	0.6	3312	1	CLR3_HUMAN
8	18	0.6	3313	1	CLR3_RAT
9	15	0.5	3298	1	PC16_HUMAN
10	13	0.4	1211	1	ATG2_HUMAN
11	13	0.4	1211	1	ATG2_HUMAN
12	13	0.4	1967	1	CD87_DROME
13	12	0.4	357	1	RLA0_METYA
14	12	0.4	1013	1	PTPX_MACNE
15	12	0.4	1015	1	PTPX_HUMAN
16	12	0.4	1121	1	MYT1_HUMAN
17	12	0.4	1310	1	ACR3_RABIT
18	12	0.4	1377	1	CBPD_MOUSE
19	11	0.4	134	1	SECR_RAT
20	11	0.4	189	1	RMP2_MOUSE
21	11	0.4	213	1	NEOM_CARAU
22	11	0.4	234	1	MDC2_HUMAN
23	11	0.4	393	1	T2D6_SCHPO
24	11	0.4	398	1	SHBG_RABIT
25	11	0.4	399	1	SHBG_PHOSU
26	11	0.4	492	1	MM11_MOUSE
27	11	0.4	527	1	GL1P_HUMAN
28	11	0.4	559	1	GL6S_CAPRI
29	11	0.4	574	1	SEN3_HUMAN
30	11	0.4	597	1	SIL1_PANTR
31	11	0.4	602	1	PGH1_RAT
32	11	0.4	617	1	ABR1_SACEX
33	11	0.4	699	1	SRCH_HUMAN

34	11	0.4	731	1	DAXX_RAT	Q9vib2	rattus norv
35	11	0.4	736	1	DAXX_CERAB	O18805	cercopithe
36	11	0.4	739	1	DAXX_MOUSE	O35613	mus musculu
37	11	0.4	740	1	DAXX_HUMAN	Q9uer7	homo sapien
38	11	0.4	741	1	PIO3_MOUSE	Q9r0e1	mus musculu
39	11	0.4	832	1	SM4B_HUMAN	Q9npr2	homo sapien
40	11	0.4	851	1	STR8_MOUSE	Q8k031	mus musculu
41	11	0.4	979	1	CLSI_MOUSE	Q9ep12	mus musculu
42	11	0.4	981	1	CLSI_HUMAN	O94985	homo sapien
43	11	0.4	1001	1	PTPX_HUMAN	P80560	mus musculu
44	11	0.4	1004	1	PTPX_RAT	O63475	rattus norv
45	11	0.4	1007	1	CHC2_HUMAN	Q9y514	homo sapien

## ALIGNMENTS

RESULT 1  
CLR2\_HUMAN STANDARD; PRT; 2923 AA.  
AC Q9HCU4; Q92566;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cadherin EGF lag seven-pass G-type receptor 2 precursor (Epidermal growth factor-like 2) (Multiple epidermal growth factor-like domains 3) (Fleming 1).  
DE 3) (Fleming 1).  
GN CELSR2 OR CDHF10 OR EGF2 OR MEGF3 OR KIAA0279.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363102; PubMed=10907856;  
RA Vincent U.B., Skang U., Scherer S.W.;  
RT "The human homologue of flamingo, EGF2, encodes a brain-expressed large cadherin-like protein with epidermal growth factor-like domains, and maps to chromosome 1p13.3-p21.1.";  
RT DNA Res. 7:233-235(2000).  
RN [2]  
RP SEQUENCE OF 516-2923 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and Brain.";  
RT DNA Res. 3:321-329(1996).  
RL - FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
CC - SUBCELLULAR LOCATION: Integral membrane protein.  
CC - TISSUE SPECIFICITY: Highest expression in brain and testis.  
CC - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC - SIMILARITY: Contains 9 cadherin domains.  
CC - SIMILARITY: Contains 8 EGF-like domains.  
CC - SIMILARITY: Contains 2 laminin G-like domains.  
CC - SIMILARITY: Contains 1 GPS domain.  
CC - This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage is for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF234887; AAC00080.1; -  
CC EMBL; D87469; BA13407.1; -  
CC HSSP; P15116; INCI.  
DR Genem; HGNC:3231; CELSR2.  
DR MIM; 604265; -

DR	GO: 0016021; C: integral to membrane; NAS.
DR	GO: 0004930; F: G-protein coupled receptor activity; NAS.
DR	InterPro; IPR000152; Axh_hydroxyl.
DR	InterPro; IPR002126; Cadherin.
DR	InterPro; IPR000742; EGF 2.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR006209; EGF-like.
DR	InterPro; IPR000832; GPCR_secretor.
DR	InterPro; IPR001879; hormu_receptor.
DR	InterPro; IPR002049; laminin_EGF.
DR	InterPro; IPR001791; Laminin_G.
DR	InterPro; IPR000203; PD_cys_rich.
DR	Pfam; PF00002; cadherin_8.
DR	Pfam; PF00008; EGF; 6.
DR	Pfam; PF01825; GPS; 1.
DR	Pfam; PF02793; HRM; 1.
DR	Pfam; PF00053; laminin_EGF_1.
DR	Pfam; PF00054; laminin_G_1.
DR	PRINTS; PR00205; CADHERIN.
DR	PRINTS; PR00015; EGF_LAMININ.
DR	PRINTS; PR00249; GPCRSECRETIN.
DR	SMART; SM00132; CA; 9.
DR	SMART; SM00179; EGF_CA; 1.
DR	SMART; SM00303; GPS; 1.
DR	SMART; SM00282; LamG; 2.
DR	PROSITE; PS00010; ASX_HYDROXYL; 2.
DR	PROSITE; PS00232; CADHERIN_1; 7.
DR	PROSITE; PS50268; CADHERIN_2; 9.
DR	PROSITE; PS50221; GPS; 1.
DR	PROSITE; PS00022; EGF_1; 6.
DR	PROSITE; PS01186; EGF_2; 4.
DR	PROSITE; PS00649; G_PROTEIN_RECP_F2_1; FALSE_NEG.
DR	PROSITE; PS00650; G_PROTEIN_RECP_F2_2; FALSE_NEG.
DR	PROSITE; PS50227; G_PROTEIN_RECP_F2_3; 1.
DR	PROSITE; PS50261; G_PROTEIN_RECP_F2_4; 1.
DR	PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR	G-protein coupled receptor; Transmembrane; Glycoprotein;
KW	EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KM	Developmental protein; Hydroxylation; Signal.
KV	
FT	SIGNAL
FT	CHAIN
FT	32
FT	2923
FT	32
FT	2380
FT	TRANSMEM
FT	2381
FT	2401
FT	2416
FT	2417
FT	2417
FT	2437
FT	2438
FT	2438
FT	2439
FT	2459
FT	2460
FT	2480
FT	2481
FT	2501
FT	2502
FT	2519
FT	2520
FT	2540
FT	2541
FT	2560
FT	TRANSMEM
FT	2561
FT	2581
FT	2582
FT	2591
FT	TRANSMEM
FT	2592
FT	2612
FT	2613
FT	2923
FT	182
FT	289
FT	290
FT	399
FT	400
FT	505
FT	DOMAIN
FT	506
FT	610
FT	712
FT	713
FT	815
FT	816
FT	921
FT	1023
FT	1028
FT	1146
FT	1228
FT	1286
FT	1288
FT	1324
FT	1328
FT	1366
FT	1571
FT	DOMAIN
FT	1367
FT	1571
FT	RECEPTOR 2.
FT	EXTRACELLULAR (POTENTIAL).
FT	1 (POTENTIAL).
FT	CYTOPLASMIC (POTENTIAL).
FT	2 (POTENTIAL).
FT	EXTRACELLULAR (POTENTIAL).
FT	3 (POTENTIAL).
FT	CYTOPLASMIC (POTENTIAL).
FT	4 (POTENTIAL).
FT	EXTRACELLULAR (POTENTIAL).
FT	5 (POTENTIAL).
FT	CYTOPLASMIC (POTENTIAL).
FT	6 (POTENTIAL).
FT	EXTRACELLULAR (POTENTIAL).
FT	7 (POTENTIAL).
FT	CYTOPLASMIC (POTENTIAL).
FT	CADHERIN 1.
FT	CADHERIN 2.
FT	CADHERIN 3.
FT	CADHERIN 4.
FT	CADHERIN 5.
FT	CADHERIN 6.
FT	CADHERIN 7.
FT	CADHERIN 8.
FT	CADHERIN 9.
FT	EGF-LIKE 1.
FT	EGF-LIKE 2.
FT	CALCIUM-BINDING.
FT	EGF-LIKE 3.
FT	CALCIUM-BINDING.
FT	LAMININ G-LIKE 1.

Query Match	Best local similarity	100.0%;	Score 2923;	DB 1;	Length 2923;			
Matches 2923;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
FT	DOMAIN	1574	1610	EGF-LIKE 4, CALCIUM-BINDING.				
FT	DOMAIN	1644	1791	EGF-LIKE 5, CALCIUM-BINDING.				
FT	DOMAIN	1793	1828	EGF-LIKE 6, CALCIUM-BINDING.				
FT	DOMAIN	1829	1867	EGF-LIKE 7, CALCIUM-BINDING.				
FT	DOMAIN	1883	1922	EGF-LIKE 8, CALCIUM-BINDING.				
FT	DOMAIN	1923	1955	EGF-LIKE 8, CALCIUM-BINDING.				
FT	DOMAIN	2316	2368	EGF-LIKE 8, CALCIUM-BINDING.				
FT	DOMAIN	2743	2752	EGF-LIKE 8, CALCIUM-BINDING.				
FT	DOMAIN	1232	1243	POLY-GLU.				
FT	DOMAIN	1237	1274	BY SIMILARITY.				
FT	DOMAIN	1276	1285	BY SIMILARITY.				
FT	DOMAIN	1292	1303	BY SIMILARITY.				
FT	DOMAIN	1297	1312	BY SIMILARITY.				
FT	DOMAIN	1314	1323	BY SIMILARITY.				
FT	DOMAIN	1332	1343	BY SIMILARITY.				
FT	DOMAIN	1337	1353	BY SIMILARITY.				
FT	DOMAIN	1355	1365	BY SIMILARITY.				
FT	DOMAIN	1578	1589	BY SIMILARITY.				
FT	DOMAIN	1583	1598	BY SIMILARITY.				
FT	DOMAIN	1600	1609	BY SIMILARITY.				
FT	DOMAIN	1797	1808	BY SIMILARITY.				
FT	DOMAIN	1802	1817	BY SIMILARITY.				
FT	DOMAIN	1819	1828	BY SIMILARITY.				
FT	DOMAIN	1832	1843	BY SIMILARITY.				
FT	DOMAIN	1837	1855	BY SIMILARITY.				
FT	DOMAIN	1857	1866	BY SIMILARITY.				
FT	DOMAIN	1887	1899	BY SIMILARITY.				
FT	DOMAIN	1889	1906	BY SIMILARITY.				
FT	DOMAIN	1908	1921	BY SIMILARITY.				
FT	DOMAIN	1924	1936	BY SIMILARITY.				
FT	DOMAIN	1926	1943	BY SIMILARITY.				
FT	DOMAIN	1945	1954	BY SIMILARITY.				
FT	DOMAIN	1951	1959	BY SIMILARITY.				
FT	DOMAIN	1810	1810	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	486	486	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	557	557	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	701	701	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1036	1036	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1076	1076	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1182	1182	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1212	1212	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1501	1501	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1565	1565	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1741	1741	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1827	1827	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1900	1900	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2024	2024	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2043	2043	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2061	2061	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2223	2223	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2345	2345	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	2923	2923	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	317447	317447	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	382757315158BD	382757315158BD	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	486	486	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	557	557	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	701	701	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1036	1036	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1076	1076	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1182	1182	HYDROXYLATION (POTENTIAL).				
FT	DOMAIN	1212						

QY VTTAEELRETKSTVFRVTLADHGMKRSALATITLVTDTNDHPVFEQOEYKESURE 300  
 Db VTTAEELRETKSTVFRVTLADHGMKRSALATITLVTDTNDHPVFEQOEYKESURE 300  
 QY NLEVEYEVLTVAATGDGAPNPANILYRLLEGSGSPSEVEFIDPRSGVIRTRGPDREY 360  
 Db NLEVEYEVLTVAATGDGAPNPANILYRLLEGSGSPSEVEFIDPRSGVIRTRGPDREY 360  
 QY ESYQUTVEASDQGRDPPGPRSTTAAVFLSYEDDNDNAPQSEKRYVVOYREBDVTPGAPYLK 420  
 Db ESYQUTVEASDQGRDPPGPRSTTAAVFLSYEDDNDNAPQSEKRYVVOYREBDVTPGAPYLK 420  
 QY VTAASRDGNSAVHYYSIMSGNARQFYLDAQTGLDVVSPLDYETTKYTLRVAQOGG 480  
 Db VTAASRDGNSAVHYYSIMSGNARQFYLDAQTGLDVVSPLDYETTKYTLRVAQOGG 480  
 QY RPLPSNBSGLVTVQVLDINDNAPIFVSPFOATVLESYPLGYLVHVOALIDADAGDNRL 540  
 Db RPLPSNBSGLVTVQVLDINDNAPIFVSPFOATVLESYPLGYLVHVOALIDADAGDNRL 540  
 QY EYRLAGVGHDPFTJINNGTGMISVAEELDREBVDYSGVEARDHGTALTSASVSYTV 600  
 Db EYRLAGVGHDPFTJINNGTGMISVAEELDREBVDYSGVEARDHGTALTSASVSYTV 600  
 QY EYRLAGVGHDPFTJINNGTGMISVAEELDREBVDYSGVEARDHGTALTSASVSYTV 600  
 Db EYRLAGVGHDPFTJINNGTGMISVAEELDREBVDYSGVEARDHGTALTSASVSYTV 600  
 QY LDVNDNNFTTPOPEYTVRLNEDAVAGTSVYVTSANDRDASHYTTQITSGNTRNFSITS 660  
 Db LDVNDNNFTTPOPEYTVRLNEDAVAGTSVYVTSANDRDASHYTTQITSGNTRNFSITS 660  
 QY LDVNDNNFTTPOPEYTVRLNEDAVAGTSVYVTSANDRDASHYTTQITSGNTRNFSITS 660  
 Db LDVNDNNFTTPOPEYTVRLNEDAVAGTSVYVTSANDRDASHYTTQITSGNTRNFSITS 660  
 QY QSGGGLVSLALPLDYKLEROYVLAATASDGTRODPAQIVVNTDANTHRPVFQSHSYTV 720  
 Db QSGGGLVSLALPLDYKLEROYVLAATASDGTRODPAQIVVNTDANTHRPVFQSHSYTV 720  
 QY VNEBDRPACTVVLISATDEDTGENARITYFMEDSIPOFRIDADGAVTTQAELEDYDQVS 780  
 Db VNEBDRPACTVVLISATDEDTGENARITYFMEDSIPOFRIDADGAVTTQAELEDYDQVS 780  
 QY YTLAATAANDNGIPQSDPTTYLEILVNDVNDNAPQRLRDSYQSGSYEDVPPFTSVQISAT 840  
 Db YTLAATAANDNGIPQSDPTTYLEILVNDVNDNAPQRLRDSYQSGSYEDVPPFTSVQISAT 840  
 QY DRDSGLNGRVFTYFGGDDGDDGFIVESTSGIVRTLRRLDRNVAQVYLRAVADKMP 900  
 Db DRDSGLNGRVFTYFGGDDGDDGFIVESTSGIVRTLRRLDRNVAQVYLRAVADKMP 900  
 QY ARTPEMEVTVTVLDVNDNPPVFEQDEPDVFEENSPIGLAVARVATDDEGNAQIMYQI 960  
 Db ARTPEMEVTVTVLDVNDNPPVFEQDEPDVFEENSPIGLAVARVATDDEGNAQIMYQI 960  
 QY VEGNIPEVFOQDIFSGELTALVDLYEDRPEYVVIQATSAVLSARATVHRLDRNDP 1020  
 Db VEGNIPEVFOQDIFSGELTALVDLYEDRPEYVVIQATSAVLSARATVHRLDRNDP 1020  
 QY PVLGAFELLFNNTYNNRSSSPFGAIGRVPADPDISLSTYSPERGNELSLVLNASTG 1080  
 Db PVLGAFELLFNNTYNNRSSSPFGAIGRVPADPDISLSTYSPERGNELSLVLNASTG 1080  
 QY ELKISRALDNNRPLEAIVSVDSVHSTVTAQCALRVTIIDEMULTHSITRLJEDMSER 1140  
 Db ELKISRALDNNRPLEAIVSVDSVHSTVTAQCALRVTIIDEMULTHSITRLJEDMSER 1140  
 QY FLSPILGFIQAVATLTATPRPHVVVFNQVQRTDAPGCHILNVSISVQCPFGGSGPPL 1200  
 Db FLSPILGFIQAVATLTATPRPHVVVFNQVQRTDAPGCHILNVSISVQCPFGGSGPPL 1200  
 QY PSEDLOERLYNRSLLTAISAQVLPFDNICTAREPCENYKCVSLRFDSSAPRTASS 1260  
 Db PSEDLOERLYNRSLLTAISAQVLPFDNICTAREPCENYKCVSLRFDSSAPRTASS 1260  
 QY VLFRIPIHVGGIARCRCPGFTGDYCYTEVDLCYSRPCGPHGCRSREGGYTCLCADGYTG 1320  
 Db VLFRIPIHVGGIARCRCPGFTGDYCYTEVDLCYSRPCGPHGCRSREGGYTCLCADGYTG 1320  
 QY BHCEVSASASGRCTPGVCNKGCTCVNLLVGGFKCDPBGDFEKPQCVQVTRSPASFITP 1380

Db BHCEVSASASGRCTPGVCNKGCTCVNLLVGGFKCDPBGDFEKPQCVQVTRSPASFITP 1380  
 QY RGLRORFFHTALFATKERDGLLYNGRPNKBDPVALVETIOEVOULTFSAGSTTVS 1440  
 Db RGLRORFFHTALFATKERDGLLYNGRPNKBDPVALVETIOEVOULTFSAGSTTVS 1440  
 QY PFVPGVSDGQMTVOLKYNNKPLLQGTGLPOGSEBOKAVAVTVQDCTGVALRGSVILG 1500  
 Db PFVPGVSDGQMTVOLKYNNKPLLQGTGLPOGSEBOKAVAVTVQDCTGVALRGSVILG 1500  
 QY NYSCAAGTQGSKKSIDLTPPLLLGVPDLPESEFPVPMROFVQCMRNLQVDSRHIDMAD 1560  
 Db NYSCAAGTQGSKKSIDLTPPLLLGVPDLPESEFPVPMROFVQCMRNLQVDSRHIDMAD 1560  
 QY FIANNGTVPCCPAKKNVCDNWTGANGGTCVQOMQAFSCBCLGFGGKSCAQBMANPOHFL 1620  
 Db FIANNGTVPCCPAKKNVCDNWTGANGGTCVQOMQAFSCBCLGFGGKSCAQBMANPOHFL 1620  
 QY GSSLVAMHGLSLPIQPMYLSIMERTROADGVLQALITRGHSTITTLQREGHVMSVBT 1680  
 Db GSSLVAMHGLSLPIQPMYLSIMERTROADGVLQALITRGHSTITTLQREGHVMSVBT 1680  
 QY GLQASSLRLEPRGRANDGMHHAQALGASGPGHAIISFDYGOORABGNLCPRLHGLHS 1740  
 Db GLQASSLRLEPRGRANDGMHHAQALGASGPGHAIISFDYGOORABGNLCPRLHGLHS 1740  
 QY NITVGGIPGPAGVARGRGCLQGVRSVDPREGNSLDPHSGESINVEQGSLLDPCDSN 1800  
 Db NITVGGIPGPAGVARGRGCLQGVRSVDPREGNSLDPHSGESINVEQGSLLDPCDSN 1800  
 QY PCPANSYCSNDMDWSYSGCDPGYVINDCTVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 Db PCPANSYCSNDMDWSYSGCDPGYVINDCTVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 QY YLGPYCETRLIDQPPRGWGHPTGCPNCNDVSKFDPDCKNTSGRCHKENHYRPPGSP 1920  
 Db YLGPYCETRLIDQPPRGWGHPTGCPNCNDVSKFDPDCKNTSGRCHKENHYRPPGSP 1920  
 QY CLLCDCTPFGSLSVCPEDPDQCPCKKGVIGRQCDRCNPPAEVYTNGBEVNYSCEPAI 1980  
 Db CLLCDCTPFGSLSVCPEDPDQCPCKKGVIGRQCDRCNPPAEVYTNGBEVNYSCEPAI 1980  
 QY BAGIMWPRTRFGLPAAPCPKGSFGIARVHRCDEHGMPLPMLFNCTSTTFSELGPAERL 2040  
 Db BAGIMWPRTRFGLPAAPCPKGSFGIARVHRCDEHGMPLPMLFNCTSTTFSELGPAERL 2040  
 QY QRNESGLDSGRSQDALLLRNATQHTAGYRGSVYKVAVOLATRLAHSTGRGGLSATQ 2100  
 Db QRNESGLDSGRSQDALLLRNATQHTAGYRGSVYKVAVOLATRLAHSTGRGGLSATQ 2100  
 QY DVHFTENILRVGSLALDTANKRHEILQOTEGGTAMLLQHYEAVASALQAMRHTVLSPF 2160  
 Db DVHFTENILRVGSLALDTANKRHEILQOTEGGTAMLLQHYEAVASALQAMRHTVLSPF 2160  
 QY TIVTENVIVSVRLDKGNFAGAKLPRYEALRGEQPDLETTVIIPEVFRETPPVBPAG 2220  
 Db TIVTENVIVSVRLDKGNFAGAKLPRYEALRGEQPDLETTVIIPEVFRETPPVBPAG 2220  
 QY PGEAQBPEBELARORRHEILSOGRAVASVIIYRLTAGLLPNNYDPDRKSLRVPKRPINT 2280  
 Db PGEAQBPEBELARORRHEILSOGRAVASVIIYRLTAGLLPNNYDPDRKSLRVPKRPINT 2280  
 QY PVASISVHDBELLPRALDKPVTVQFRLLEBRTKICVPMNHSILVSGTGSAGSAGE 2340  
 Db PVASISVHDBELLPRALDKPVTVQFRLLEBRTKICVPMNHSILVSGTGSAGSAGE 2340  
 QY VVFNESHVSQCNHMTSPAVLNDVSRRENGEILPLKTLTVVALGVTLAALLTFPFLTL 2400  
 Db VVFNESHVSQCNHMTSPAVLNDVSRRENGEILPLKTLTVVALGVTLAALLTFPFLTL 2400  
 QY LRIIRSNQHRBMLTALGIAOLVPLGINQDMLPACTVIALLLFLVICTESMLLE 2460



```

Db 2401 LRLRSNONGIRRNLTAAQLAQLVLLGINQADLPFACTIVAILLFLYLCTFSWALLE 2460
Qy 2461 ALHLRYALTEVRDVNTGPMKFTYMLGNGVPAFTGLAVGLDPKGNPFCMTSTYDTLI 2520
Db 2461 ALHLRYALTEVRDVNTGPMKFTYMLGNGVPAFTGLAVGLDPKGNPFCMTSTYDTLI 2520
Qy 2521 WSPGAVPAFVNSVFLYTLAARASCAHQGFEEKGPGVGLQPSRAVLLLSATWLLAL 2580
Db 2521 WSPGAVPAFVNSVFLYTLAARASCAHQGFEEKGPGVGLQPSRAVLLLSATWLLAL 2580
Qy 2581 LSVNSTLLPHYLPAFCNCTCGPFILSYVLSKEYRAKALKACSKRPSPDPALTTKSTL 2640
Db 2581 LSVNSTLLPHYLPAFCNCTCGPFILSYVLSKEYRAKALKACSKRPSPDPALTTKSTL 2640
Qy 2641 TSSYNCPSPYADRLQYVYDSDAGSLHSTSRSGKSPSYPLLRBSALNPGQPGG 2700
Db 2641 TSSYNCPSPYADRLQYVYDSDAGSLHSTSRSGKSPSYPLLRBSALNPGQPGG 2700
Qy 2701 DPGSLFLBQDQDHPDPTSDLSLEDDQSGSYASTHSSDSSEBEEBEEBAAPFGEQ 2760
Db 2701 DPGSLFLBQDQDHPDPTSDLSLEDDQSGSYASTHSSDSSEBEEBEEBAAPFGEQ 2760
Qy 2761 WBSLGPQARLPPLHSTPQDGGPGKAPPPGDTTAKSSGNGAPBERLRNPDALSR 2820
Db 2761 WBSLGPQARLPPLHSTPQDGGPGKAPPPGDTTAKSSGNGAPBERLRNPDALSR 2820
Qy 2821 EGSGLGPLPGSSAQPHHGILKKCLPTISEKSLRLPLBQCTSSSGSSASBSRGGPP 2880
Db 2821 EGSGLGPLPGSSAQPHHGILKKCLPTISEKSLRLPLBQCTSSSGSSASBSRGGPP 2880
Qy 2881 RPPPRQSLQEQNLGVNPIAMISIRACTVDEDSGSEFLFFNPLH 2923
Db 2881 RPPPRQSLQEQNLGVNPIAMISIRACTVDEDSGSEFLFFNPLH 2923

RESULT 2
CLR2 RAT STANDARD; PRT; 2144 AA.
AC 090YF2;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caderlin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal
   growth factor-like domains 3) (Fragment).
GN CBLSR2 OR MEGF3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
   EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
   signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. High expression in
   cerebellum and olfactory bulb. Weaker expression in cerebral
   cortex, hippocampus and brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 4 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011529; BAA8687.1; -.
DR HSSP; P00740; 1EDM.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0004930; P:G-protein coupled receptor activity; ISS.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00008; cadherin; 3.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00053; laminin_EGF; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGPLAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 3.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Homr; 1.
DR SMART; SM00282; Lamc; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS02068; CADHERIN_2; 4.
DR PROSITE; PS50221; GPS; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE; PS50227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPR_EGF; 1.
KM G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation.
FT FT 1 1605 EXTRACELLULAR (POTENTIAL).
FT FT 1 1626 1 (POTENTIAL).
FT FT 1627 1641 CYTOPLASMIC (POTENTIAL).
FT FT 1642 1662 2 (POTENTIAL).
FT FT 1663 1683 EXTRACELLULAR (POTENTIAL).
FT FT 1684 1694 3 (POTENTIAL).
FT FT 1685 1705 CYTOPLASMIC (POTENTIAL).
FT FT 1706 1726 4 (POTENTIAL).
FT FT 1727 1744 EXTRACELLULAR (POTENTIAL).
FT FT 1745 1765 5 (POTENTIAL).
FT FT 1766 1789 CYTOPLASMIC (POTENTIAL).
FT FT 1790 1810 6 (POTENTIAL).
FT FT 1811 1816 EXTRACELLULAR (POTENTIAL).
FT FT 1817 1837 7 (POTENTIAL).
FT FT 1838 2144 CYTOPLASMIC (POTENTIAL).
FT FT <1 40 CADHERIN 1.
FT FT 41 146 CADHERIN 2.
FT FT 147 248 CADHERIN 3.
FT FT 253 371 CADHERIN 4.
FT FT 453 511 EGF-LIKE 1.
FT FT 513 549 EGF-LIKE 2.
FT FT 553 591 EGF-LIKE 3.
FT FT 592 796 LAMININ G-LIKE 1.

```

```

FT DOMAIN 799 835 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 839 1016 LAMININ G-LIKE 2.
FT DOMAIN 1018 1053 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 1054 1092 EGF-LIKE 6, CALCIUM-BINDING.
FT DOMAIN 1108 1147 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 1148 1180 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 1541 1593 GPG.
FT DOMAIN 1968 1973 POLY-GLU.
FT DISULFID 457 468 BY SIMILARITY.
FT DISULFID 462 499 BY SIMILARITY.
FT DISULFID 501 510 BY SIMILARITY.
FT DISULFID 517 528 BY SIMILARITY.
FT DISULFID 522 537 BY SIMILARITY.
FT DISULFID 539 548 BY SIMILARITY.
FT DISULFID 557 568 BY SIMILARITY.
FT DISULFID 562 578 BY SIMILARITY.
FT DISULFID 580 590 BY SIMILARITY.
FT DISULFID 803 814 BY SIMILARITY.
FT DISULFID 808 823 BY SIMILARITY.
FT DISULFID 825 834 BY SIMILARITY.
FT DISULFID 1022 1033 BY SIMILARITY.
FT DISULFID 1027 1042 BY SIMILARITY.
FT DISULFID 1044 1053 BY SIMILARITY.
FT DISULFID 1057 1068 BY SIMILARITY.
FT DISULFID 1062 1080 BY SIMILARITY.
FT DISULFID 1082 1091 BY SIMILARITY.
FT DISULFID 1112 1124 BY SIMILARITY.
FT DISULFID 1113 1131 BY SIMILARITY.
FT DISULFID 1149 1161 BY SIMILARITY.
FT DISULFID 1151 1168 BY SIMILARITY.
FT DISULFID 1170 1179 BY SIMILARITY.
FT MOD_RES 816 816 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1035 1035 HYDROXYLATION (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 966 966 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1249 1249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1286 1286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1548 1548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1570 1570 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;

```

```

Query Match 6.4%; Score 187; DB 1; Length 2144;
Best Local Similarity 100.0%; Pred. No. 3e-134;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1115 LKVTITITDMLTHSTLRLEDMSPRRPLSLGLFTQNAATLAPPPDHVVVYVNNQRD 1174
DB 340 LKVTITITDMLTHSTLRLEDMSPRRPLSLGLFTQNAATLAPPPDHVVVYVNNQRD 399
QY 1175 ABGCHILNLSVSGVPPGPGGPPFLPSEDLQERLYLNLSLTALISAQVLPFDNICLR 1234
DB 400 ABGCHILNLSVSGVPPGPGGPPFLPSEDLQERLYLNLSLTALISAQVLPFDNICLR 459
QY 1235 BECEYMCVSVLRPDSAPFIASSSVLFRPIHPVGGLRCRCPGFTGDCYCEVDLCYS 1294
DB 460 BECEYMCVSVLRPDSAPFIASSSVLFRPIHPVGGLRCRCPGFTGDCYCEVDLCYS 519
QY 1295 RPCGPHG 1301
DB 520 RPCGPHG 526

```

```

ID CLAR MOUSE STANDARD; PRT; 2920 AA.
AC O9R0M0; Q99K26; Q92ZK4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
GN (mp11).
OS CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed=10490098;
RX MEDLINE=99418630; PubMed=10490098;
RA Ueno T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled."
RL Cell 98:585-595(1999).
RN [2]
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20253755; PubMed=10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.P.R.;
RT "Chromosomal localization of Celser2 and Celser3 in the mouse; Celser3 is
RT a candidate for the lipply (ltp) lethal mutant on chromosome 9."
RL Mamm. Genome 11:392-394(2000).
RN [3]
RP SEQUENCE OF 2014-2920 FROM N.A.
RX TISSUE=Breast tumor;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Reinhold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marnulna K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP DEVELOPMENTAL STAGE.
RX MEDLINE=21839555; PubMed=11850187;
RA Tissier F., De-Becker O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celser (Flamingo) genes in the
RT mouse."
RL Mech. Dev. 112:157-160(2002).
RN [5]
RP FUNCTION: Receptor that may have an important role in cell/cell
RN signaling during nervous system formation.
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
CC -1- DEVELOPMENTAL STAGE: Predominantly expressed in the developing
CC CNS, the emerging dorsal root ganglia and cranial ganglia. In the
CC CNS, expression is uniform along the rostrocaudal axis. During
CC gastrulation, it is expressed within the anterior neural ectoderm.
CC At E10, expression is strong in the ventricular zones (VZ) in all
CC sectors of the brain, and lower in the marginal zones (MZ).
CC Between E12 and E15, expression is prominent in the brain. It is
CC strong in VZ, lower in MZ, except in telecephalic MZ where it is
CC predominant. The intensity is higher in all VZ, and lower in
CC differentiating fields than in VZ, except in the cerebral
CC hemispheres, and to a lesser extent in the tectum and cerebellum.

```

RESULT 3  
CLAR\_MOUSE

CC A weak expression is also observed in the fetal lungs, kidney and  
CC epithelia. In the newborn and postnatal stages, expression remains  
CC restricted to the VZ as well as in migrating and postmigratory  
CC cells throughout the brain.

CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC  
CC -1- SIMILARITY: Contains 9 cadherin domains.  
CC  
CC -1- SIMILARITY: Contains 8 EGF-like domains.  
CC  
CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
CC  
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.  
CC  
CC -1- SIMILARITY: Contains 1 Gps domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMLL; AB028499; BAA84070.1; -
DR	EMLL; AF031573; AAC68837.1; -
DR	EMLL; BC005499; AAH05499.1; -
DR	HSSP; P00740; IEDM.
DR	MGD; MGI:1856235; Celsr2.
DR	GO; GO:0016021; C:integral to membrane; ISS.
DR	GO; GO:0049930; F:g-protein coupled receptor activity; ISS.
DR	InterPro; IPRO00152; Asx hydroxyl.
DR	InterPro; IPRO02126; Cadherin.
DR	InterPro; IPRO00742; EGF_2.
DR	InterPro; IPRO06209; EGF_like.
DR	InterPro; IPRO00832; GPCR_secretin.
DR	InterPro; IPRO02849; hormn_receptor.
DR	InterPro; IPRO01791; Laminin_G.
DR	InterPro; IPRO00203; PKD_cys_rich.
DR	Pfam; PF00002; Tm_2; 1.
DR	Pfam; PF00028; cadherin; 9.
DR	Pfam; PF00008; EGF; 5.
DR	Pfam; PF01825; GPS; 1.
DR	Pfam; PF02793; HRM; 1.
DR	Pfam; PF00053; laminin_EGF; 1.
DR	Pfam; PF00054; laminin_G; 2.
DR	PRINTS; PR00205; CADHERIN.
DR	PRINTS; PRO0011; EGFLAMININ.
DR	PRINTS; PRO0249; GPORSECRETIN.
DR	SMART; SMO0112; CA; 9.
DR	SMART; SMO0180; EGF_Lam; 1.
DR	SMART; SMO0303; GPS; 1.
DR	SMART; SMO0008; HORMR; 1.
DR	SMART; SMO0282; Lamg; 2.
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.
DR	PROSITE; PS00232; CADHERIN_1; 6.
DR	PROSITE; PS02668; CADHERIN_2; 9.
DR	PROSITE; PS00022; EGF_1; 6.
DR	PROSITE; PS01866; EGF_2; 2.
DR	PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR	PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR	PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR	PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR	PROSITE; PS02221; GPS; 1.
DR	PROSITE; PS00025; IAM_G_DOMAIN; 2.
KW	G-protein coupled receptor; Transmembrane; Glycoprotein;
KW	EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW	Developmental protein; Hydroxylation; Signal.
FT	SIGNAL
FT	CHAIN
FT	32
FT	2920
FT	32
FT	2381
FT	2382
FT	2402
FT	DOMAIN
FT	2403
FT	2414
FT	TRANSMEM
FT	2415
FT	2434
FT	2435
FT	2439
FT	DOMAIN
FT	2435
FT	2439
FT	EXTRACELLULAR (POTENTIAL).
FT	CADHERIN EGF LAG SEVEN-PASS G-TYPE RECEPTOR 2.
FT	EXTRACELLULAR (POTENTIAL).
FT	1 (POTENTIAL).
FT	CYTOSOLASMIC (POTENTIAL).
FT	2 (POTENTIAL).
FT	EXTRACELLULAR (POTENTIAL).

FT	TRANSSEM	2440	2460	3 (POTENTIAL)
FT	DOMAIN	2461	2481	CYTOPLASMIC (POTENTIAL) .
FT	TRANSSEM	2482	2502	4 (POTENTIAL) .
FT	DOMAIN	2503	2519	EXTRACELLULAR (POTENTIAL) .
FT	TRANSSEM	2520	2540	5 (POTENTIAL) .
FT	DOMAIN	2541	2564	CYTOPLASMIC (POTENTIAL) .
FT	TRANSSEM	2565	2585	6 (POTENTIAL) .
FT	DOMAIN	2586	2592	EXTRACELLULAR (POTENTIAL) .
FT	TRANSSEM	2593	2613	7 (POTENTIAL) .
FT	DOMAIN	2614	2920	CYTOPLASMIC (POTENTIAL) .
FT	DOMAIN	182	289	CADHERIN 1 .
FT	DOMAIN	290	399	CADHERIN 2 .
FT	DOMAIN	400	506	CADHERIN 3 .
FT	DOMAIN	507	611	CADHERIN 4 .
FT	DOMAIN	612	713	CADHERIN 5 .
FT	DOMAIN	714	816	CADHERIN 6 .
FT	DOMAIN	817	922	CADHERIN 7 .
FT	DOMAIN	923	1024	CADHERIN 8 .
FT	DOMAIN	1029	1147	CADHERIN 9 .
FT	DOMAIN	1229	1287	EGF-LIKE 1 (ATYPICAL) .
FT	DOMAIN	1289	1319	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	1329	1367	EGF-LIKE 3, CALCIUM-BINDING.
FT	DOMAIN	1368	1572	LAMININ G-LIKE 1 .
FT	DOMAIN	1575	1611	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	1615	1792	LAMININ G-LIKE 2 .
FT	DOMAIN	1788	1830	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1831	1868	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	1924	1923	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	1924	1956	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	2317	2369	GPS .
FT	DOMAIN	2744	2749	POLY-GU .
FT	DISULFID	1293	1304	BY SIMILARITY .
FT	DISULFID	1298	1313	BY SIMILARITY .
FT	DISULFID	1315	1318	BY SIMILARITY .
FT	DISULFID	1333	1344	BY SIMILARITY .
FT	DISULFID	1338	1354	BY SIMILARITY .
FT	DISULFID	1356	1366	BY SIMILARITY .
FT	DISULFID	1579	1590	BY SIMILARITY .
FT	DISULFID	1584	1599	BY SIMILARITY .
FT	DISULFID	1601	1610	BY SIMILARITY .
FT	DISULFID	1792	1803	BY SIMILARITY .
FT	DISULFID	1798	1818	BY SIMILARITY .
FT	DISULFID	1820	1829	BY SIMILARITY .
FT	DISULFID	1833	1844	BY SIMILARITY .
FT	DISULFID	1838	1856	BY SIMILARITY .
FT	DISULFID	1858	1867	BY SIMILARITY .
FT	DISULFID	1888	1900	BY SIMILARITY .
FT	DISULFID	1890	1907	BY SIMILARITY .
FT	DISULFID	1909	1922	BY SIMILARITY .
FT	DISULFID	1925	1937	BY SIMILARITY .
FT	DISULFID	1927	1944	BY SIMILARITY .
FT	DISULFID	1946	1955	BY SIMILARITY .
FT	MOD_RES	1592	1592	HYDROXYLATION (POTENTIAL) .
FT	CARBONYL	486	486	N-LINKED (GLCNAC . .) (POTENTIAL) .

Query Match	5.7%	Score 168	DB 1	Length 2920
Best Local Similarity	100.0%	Pred. No. 1.2e-119		
Matches 168	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	915	NDNPVFEQDEPFVFEHNSPIGLAVARVATATPDDEGTNAQIMQIVEGNIPEVFEOLDIF	974	
Db	916	NDNPVFEQDEPFVFEHNSPIGLAVARVATATPDDEGTNAQIMQIVEGNIPEVFEOLDIF	975	
QY	975	SGELTALVDDYEDRPEYVLVIQATSAPLVSRAIVHRLDRNDNPVLGNFELLFNNYV	1034	
Db	976	SGELTALVDDYEDRPEYVLVIQATSAPLVSRAIVHRLDRNDNPVLGNFELLFNNYV	1035	
QY	1035	TNNSSSPPGGAIGRPVAPHDPIDSLTSPFERGNEISLVILNASTGEL	1082	
Db	1036	TNNSSSPPGGAIGRPVAPHDPIDSLTSPFERGNEISLVILNASTGEL	1083	
RESULT 4				

CLRL_MOUSE	STANDARD:	PRT;	3034 AA.
ID	CLRL_MOUSE		
AC	O35161;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Cadherin EGF LAG seven-pass G-type receptor 1 precursor.		
GN	CELSR1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
NCBI	NCBI_Txid=10090;		
XX	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99077689; PubMed=9658697;		
RA	Hadjantonakis A.-K., Formstone C.J., Little P.F.R.;		
RT	"mcelr1 is an evolutionarily conserved seven-pass transmembrane receptor and is expressed during mouse embryonic development.";		
RL	Mech. Dev. 78:91-95(1998).		
RL	[2]		
RP	TISSUE SPECIFICITY.		
RC	STRAIN=C57BL/6; TISSUE=Brain;		
RX	MEDLINE=97480720; PubMed=9339365;		
RA	Hadjantonakis A.-K., Sheward W.J., Hartmar A.J., de Galan L.,		
RT	Hoover J.M.N., Little P.F.R.;		
RT	"Celsr1, a neural-specific gene encoding an unusual seven-pass transmembrane receptor, maps to mouse chromosome 15 and human chromosome 22qter.";		
RT	Genomics 45:97-104(1997).		
RL	[3]		
RP	DEVELOPMENTAL STAGE.		
RX	MEDLINE=21839555; PubMed=11850187;		
RA	Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;		
RT	"Developmental expression profiles of Celsr (Flamingo) genes in the mouse.";		
RL	Mech. Dev. 112:157-160(2002).		
CC	-1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	-1- TISSUE SPECIFICITY: Expressed in the brain, where it is localized principally in the ependymal cell layer, choroid plexus and the area postrema. Also found in spinal chord and in the eye.		
CC	-1- DEVELOPMENTAL STAGE: First detected at E6. Predominantly expressed in the developing CNS, the emerging dorsal root ganglia and cranial ganglia. In the CNS, expression is uniform along the rostrocaudal axis. During gastrulation, it is expressed in the vicinity of the primitive streak, and becomes predominant in that area at late gastrulation. At E10, detected in ventricular zones (VZ), but not in marginal zones (MZ), and weakly in other structures. Between E12 and E15, a high expression is present in the VZ in all brain areas. No expression in differentiated neuronal fields. In the newborn and postnatal stages, expression remains restricted to the VZ. Also found weakly in fetal lungs, kidney and epithelia.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.		
CC	-1- SIMILARITY: Contains 9 cadherin domains.		
CC	-1- SIMILARITY: Contains 8 EGF-like domains.		
CC	-1- SIMILARITY: Contains 2 laminin G-like domains.		
CC	-1- SIMILARITY: Contains 1 laminin EGF-like domain.		
CC	-1- SIMILARITY: Contains 1 GPS domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and that this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; AF031572; AAC68836.1; ..		
DR	PIR; T14119; T14115.		
DR	MGD; MGI:1100883; Celsr1.		
DR	HSSP; P00749; IURK.		

DR	InterPro: IPR000152; Asx_hydroxyl.
DR	InterPro: IPR002126; Cadherin.
DR	InterPro: IPR006209; EGF-like.
DR	InterPro: IPR000832; GPCR secretin.
DR	InterPro: IPR001879; homin_receptor.
DR	InterPro: IPR002049; laminin_EGF.
DR	InterPro: IPR001791; laminin_G.
DR	InterPro: IPR000203; PKD_cys_rich.
DR	Pfam: PF00002; 7tm_2; 1. cys_rich.
DR	Pfam: PF00028; cadherin. 9.
DR	Pfam: PR00008; EGF; 6.
DR	Pfam: PF01825; GFS; 1.
DR	Pfam: PF02793; HRM; 1.
DR	Pfam: PF00053; laminin_EGF; 1.
DR	Pfam: PF00054; laminin_G; 1.
DR	PRINTS: PR00249; CADHERIN.
DR	PRINTS: PR00011; EGPLAMININ.
DR	PRINTS: PR00249; GPCSECRETIN.
DR	SMART: SM00112; CA; 9.
DR	SMART: SM00180; EGF_Lam; 1.
DR	SMART: SM00303; GFS; 1.
DR	SMART: SM00008; Homr; 1.
DR	SMART: SM00282; Lamc; 2.
DR	PROSITE: PS00010; ASX_HYDROXYL; 2.
DR	PROSITE: PS00232; CADHERIN_1; 7.
DR	PROSITE: PS0268; CADHERIN_2; 9.
DR	PROSITE: PS00022; EGF_1; 6.
DR	PROSITE: PS0186; EGF_2; 2.
DR	PROSITE: PS0221; GFS; 1.
DR	PROSITE: PS00649; G_PROTEIN_RECP_F2_1; FALSE NEG.
DR	PROSITE: PS00650; G_PROTEIN_RECP_F2_2; FALSE NEG.
DR	PROSITE: PS00227; G_PROTEIN_RECP_F2_3; 1.
DR	PROSITE: PS0261; G_PROTEIN_RECP_F2_4; 1.
DR	PROSITE: PS01248; LAMININ_TYR_EGF_1.
DR	PROSITE: PS0025; LAM_G_DOMAIN; 2.
KW	G-protein coupled receptor; Transmembrane; Glycoprotein;
KW	EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;
KW	Developmental protein; Hydroxylation; Signal.
KM	SIGNAL
FT	CHAIN 1 29
FT	CHAIN 30 3034
FT	DOMAIN 21 2484
FT	TRANSMEM 2485 2505
FT	DOMAIN 2506 2516
FT	TRANSMEM 2517 2537
FT	DOMAIN 2538 2542
FT	TRANSMEM 2543 2563
FT	DOMAIN 2564 2587
FT	TRANSMEM 2588 2608
FT	DOMAIN 2609 2625
FT	TRANSMEM 2626 2646
FT	DOMAIN 2647 2670
FT	TRANSMEM 2671 2691
FT	DOMAIN 2692 2694
FT	TRANSMEM 2695 2715
FT	DOMAIN 2716 3034
FT	DOMAIN 251 368
FT	DOMAIN 369 474
FT	DOMAIN 475 580
FT	DOMAIN 581 702
FT	DOMAIN 703 804
FT	DOMAIN 805 907
FT	DOMAIN 908 1014
FT	DOMAIN 1015 1116
FT	DOMAIN 1121 1239
FT	DOMAIN 1318 1376
FT	DOMAIN 1378 1414
FT	DOMAIN 1418 1456
FT	DOMAIN 1457 1661
FT	DOMAIN 1664 1700
FT	DOMAIN 1704 1885
FT	DOMAIN 1887 1922
FT	DOMAIN 1923 1961

```

FT DOMAIN 1962 1994 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 1996 2031 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 2037 2070 LAMININ EGF-LIKE.
FT DOMAIN 2423 2475 GFS.
FT DOMAIN 2674 2678 POLY-LEU.
FT DISULFID 1322 1333 BY SIMILARITY.
FT DISULFID 1327 1364 BY SIMILARITY.
FT DISULFID 1366 1375 BY SIMILARITY.
FT DISULFID 1382 1393 BY SIMILARITY.
FT DISULFID 1387 1402 BY SIMILARITY.
FT DISULFID 1404 1413 BY SIMILARITY.
FT DISULFID 1422 1433 BY SIMILARITY.
FT DISULFID 1427 1443 BY SIMILARITY.
FT DISULFID 1445 1455 BY SIMILARITY.
FT DISULFID 1668 1679 BY SIMILARITY.
FT DISULFID 1673 1699 BY SIMILARITY.
FT DISULFID 1690 1699 BY SIMILARITY.
FT DISULFID 1891 1902 BY SIMILARITY.
FT DISULFID 1896 1911 BY SIMILARITY.
FT DISULFID 1913 1922 BY SIMILARITY.
FT DISULFID 1926 1937 BY SIMILARITY.
FT DISULFID 1931 1949 BY SIMILARITY.
FT DISULFID 1951 1960 BY SIMILARITY.
FT DISULFID 1968 1981 BY SIMILARITY.
FT DISULFID 1983 1993 BY SIMILARITY.
FT DISULFID 2000 2015 BY SIMILARITY.
FT DISULFID 2002 2018 BY SIMILARITY.
FT DISULFID 2020 2030 BY SIMILARITY.
FT MOD_RES 1681 1681 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1904 1904 HYDROXYLATION (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 793 793 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1129 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1228 1228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1264 1264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1591 1591 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1655 1655 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1994 1994 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2118 2118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2137 2137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2144 2144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2155 2155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2160 2160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2452 2452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2538 2538 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3034 AA; 330477 MW; EEPF38180AF5ED8A8 CRC64;

Query Match 0.8%; Score 23; DB 1; Length 3034;
Best Local Similarity 100.0%; Pred. No. 4,4e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

GN CELSR1 OR CDHP9 OR FM12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
RT feature of protocadherin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Brunkiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Amescough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.P., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.B., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.W., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Levergha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavlit R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R., Ramsay H., Ramsey Y., Rogers L., Rose M.T.,
RA Scott C.B., Shira H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilner T.E., Wilting L.N.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshiaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fan Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Lon P., Malat E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaul S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saita S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow J.B., Edelmann L.,
RA Kim U.J., Shitaya H., Simon M.I., Dumaneki J.P., Peyraud M., Kedra D.,
RA Seroussi E., Fraenkel I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SEQUENCE OF 624-3014 FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrincci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

RA Fahney J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
 RA Buttrickfield Y.S.N., Krzyzanski M.I., Skalka U., Smalhus D.E.,  
 RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Receptor that may have an important role in cell/cell  
 CC signaling during nervous system formation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9NYQ6-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9NYQ6-2; Sequence=VSP\_002011, VSP\_002012;  
 CC Note=No experimental confirmation available;  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: Contains 9 cadherin domains.  
 CC -1- SIMILARITY: Contains 8 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF231024; AAF61930.1; -;  
 DR EMBL, AL021392; CAB50207.1; -;  
 DR EMBL, AL031597; CAB45020.1; ALT\_INT.  
 DR EMBL, AL031588; CAB38413.1; -;  
 DR EMBL, BC000059; AAH00059.1; -;  
 DR GeneW, HGNC:1850; CELSR1.  
 DR MIM, 604523; -;  
 DR HSSP, P00749; IURK.  
 DR GO, GO:0016021, C:integral to membrane; NAS.  
 DR GO, GO:0008014, F:calcium-dependent cell adhesion molecule ac. .; NAS.  
 DR GO, GO:0004930, F:G-protein coupled receptor activity; NAS.  
 DR GO, GO:0007399, P:neurogenesis; NAS.  
 DR InterPro, IPR000152; Asx hydroxyl.  
 DR InterPro, IPR002126; Cadherin.  
 DR InterPro, IPR006209; EGF like.  
 DR InterPro, IPR000832; GPCR\_secretin.  
 DR InterPro, IPR001879; horum\_receptor.  
 DR InterPro, IPR002049; laminin\_EGF.  
 DR InterPro, IPR001791; laminin\_G.  
 DR InterPro, IPR000203; PKD\_cys\_rich.  
 DR Pfam, PF00002; 7tm\_2; 1.  
 DR Pfam, PF00028; cadherin; 8.  
 DR Pfam, PF00008; EGF; 6.  
 DR Pfam, PF01825; GPS; 1.  
 DR Pfam, PF02793; HRM; 1.  
 DR Pfam, PF00053; laminin\_EGF; 1.  
 DR Pfam, PF00054; laminin\_G; 1.  
 DR PRINTS, PR00205; CADHERIN.  
 DR PRINTS, PR00011; EGF\_LAMININ.  
 DR PRINTS, PR00249; GPCRSECRETIN.  
 DR SMART, SM00112; CA; 8.  
 DR SMART, SM00180; EGF\_Lam; 1.  
 DR SMART, SM00303; GPS; 1.  
 DR SMART, SM00008; Horum; 1.  
 DR SMART, SM00282; Lamg; 2.  
 DR PROSITE, PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE, PS00232; CADHERIN\_1; 7.  
 DR PROSITE, PS00268; CADHERIN\_2; 9.  
 DR PROSITE, PS00022; EGF\_1; 6.

DR PROSITE, PS01186; EGF\_2; 2.  
 DR PROSITE, PS00221; GPS; 1.  
 DR PROSITE, PS00649; G PROTEIN RECP F2\_1; FALSE NEG.  
 DR PROSITE, PS00650; G PROTEIN RECP F2\_2; FALSE NEG.  
 DR PROSITE, PS00227; G PROTEIN RECP F2\_3; 1.  
 DR PROSITE, PS00261; G PROTEIN RECP F2\_4; 1.  
 DR PROSITE, PS01248; LAMININ\_EGF; 1.  
 DR PROSITE, PS00025; LAM\_G\_DOMAIN; 2.  
 DR EGF-protein coupled receptor; Transmembrane; Glycoprotein;  
 DR EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;  
 DR Developmental protein; Hydroxylation; Signal; Alternative splicing.  
 FT CHAIN 1 20  
 FT DOMAIN 21 3014  
 FT TRANSMEM 22 2469  
 FT TRANSMEM 2470 2490  
 FT TRANSMEM 2491 2501  
 FT TRANSMEM 2502 2522  
 FT TRANSMEM 2523 2527  
 FT TRANSMEM 2528 2548  
 FT TRANSMEM 2549 2572  
 FT TRANSMEM 2573 2593  
 FT TRANSMEM 2594 2611  
 FT TRANSMEM 2612 2632  
 FT TRANSMEM 2633 2655  
 FT TRANSMEM 2656 2676  
 FT TRANSMEM 2677 2683  
 FT TRANSMEM 2684 2704  
 FT TRANSMEM 2705 3014  
 FT DOMAIN 246 353  
 FT DOMAIN 354 459  
 FT DOMAIN 460 565  
 FT DOMAIN 566 687  
 FT DOMAIN 688 789  
 FT DOMAIN 790 892  
 FT DOMAIN 893 999  
 FT DOMAIN 1000 1101  
 FT DOMAIN 1106 1224  
 FT DOMAIN 1303 1361  
 FT DOMAIN 1363 1399  
 FT DOMAIN 1403 1441  
 FT DOMAIN 1442 1646  
 FT DOMAIN 1649 1685  
 FT DOMAIN 1689 1870  
 FT DOMAIN 1872 1907  
 FT DOMAIN 1908 1946  
 FT DOMAIN 1947 1979  
 Query Match 0.7%; Score 20; DB 1; Length 3014;  
 Best Local Similarity 100.0%; Pred. No. 8,4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1222 QVLPEDNLCRPECNYM 1241  
 DB 1297 QVLPEDNLCRPECNYM 1316  
 RESULT 6  
 CTR3 MOUSE STANDARD; PRT; 3301 AA.  
 AC Q91T10; Q9BS00;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor.  
 GN CELSR3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.  
 RC STRAIN=C57BL/6;





```

FT DISULFID 1944 1959 BY SIMILARITY.
FT DISULFID 1961 1970 BY SIMILARITY.
FT DISULFID 1974 1985 BY SIMILARITY.
FT DISULFID 1997 1997 BY SIMILARITY.
FT DISULFID 1999 2008 BY SIMILARITY.
FT DISULFID 2016 2029 BY SIMILARITY.
FT DISULFID 2031 2041 BY SIMILARITY.
FT DISULFID 2048 2063 BY SIMILARITY.
FT DISULFID 2050 2066 BY SIMILARITY.
FT DISULFID 2068 2078 BY SIMILARITY.
FT MOD RES 1952 1952 HYDROXYLATION (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 2713 2713 L -> LR (IN REF. 2).
FT CONFLICT 3024 3024 R -> P (IN REF. 2).
SQ SEQUENCE 3301 AA; 358455 MW; A6B18F2DF7F4DEB6 CRC64;

```

```

Query Match 0.6%; Score 18; DB 1; Length 3301;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2495 GLAVGLDPGEGYGNPDFCM 2512
Db 2649 GLAVGLDPGEGYGNPDFCM 2666

```

```

RESULT 7
CLR3 HUMAN STANDARD; PRT; 3312 AA.
AC 09NY07; 075092;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo
DE homolog 1) (hFml1) (Multiple epidermal growth factor-like domains 2)
DE (Epidermal growth factor-like 1).
GN CELSR3 OR CDHP11 OR FMI1 OR EGF11 OR MEGF2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCB1_TaxId=9606;
RX 11
RP SEQUENCE FROM N.A.
RA MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
RT feature of protocadherin genes";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 1954-3312 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9836089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

```

```

CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF231023; AAF61929.1; -.
DR EMBL; AB011536; BAA32464.1; -.
DR HSSP; P00740; IEDM.
DR Genew; HGNC:3230; CELSR3.
DR MIM; 604264; -.
DR GO; GO:0005198; F: structural molecule activity; NAS.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormu_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR001791; laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G_2.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormu; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS00022; EGF_2; 4.
DR PROSITE; PS01186; GPS; 1.
DR PROSITE; PS01186; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE NEG.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS00261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS00025; LAM G DOMAIN; 2.
DR PROSITE; PS01248; LAMININ TYPE EGF; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 32
FT CHAIN 33 3312
FT DOMAIN 33 2540
FT TRANSMEM 2541 2561
FT DOMAIN 2562 2572
FT TRANSMEM 2573 2593
FT DOMAIN 2594 2601
FT TRANSMEM 2602 2622
FT DOMAIN 2623 2643
FT TRANSMEM 2644 2664
FT DOMAIN 2665 2681
FT TRANSMEM 2682 2702

```

```

PT DOMAIN 2703 2725 CYTOPLASMIC (POTENTIAL).
PT TRANSSEM 2726 2746 6 (POTENTIAL).
PT DOMAIN 2747 2753 EXTRACELLULAR (POTENTIAL).
PT TRANSSEM 2754 2774 7 (POTENTIAL).
PT DOMAIN 2775 3312 CYTOPLASMIC (POTENTIAL).
PT DOMAIN 326 433 CADHERIN 1.
PT DOMAIN 434 545 CADHERIN 2.
PT DOMAIN 546 651 CADHERIN 3.
PT DOMAIN 652 756 CADHERIN 4.
PT DOMAIN 757 858 CADHERIN 5.
PT DOMAIN 859 961 CADHERIN 6.
PT DOMAIN 962 1067 CADHERIN 7.
PT DOMAIN 1068 1169 CADHERIN 8.
PT DOMAIN 1170 1265 CADHERIN 9.
PT DOMAIN 1266 1433 EGF-LIKE 1.
PT DOMAIN 1434 1471 EGF-LIKE 2.
PT DOMAIN 1472 1514 EGF-LIKE 3.
PT DOMAIN 1515 1719 EGF-LIKE 4.
PT DOMAIN 1720 1758 EGF-LIKE 5.
PT DOMAIN 1759 1944 EGF-LIKE 6.
PT DOMAIN 1945 1982 EGF-LIKE 7.
PT DOMAIN 1983 2020 EGF-LIKE 8.
PT DOMAIN 2021 2053 EGF-LIKE 9.
PT DOMAIN 2054 2090 EGF-LIKE 10.
PT DOMAIN 2091 2131 EGF-LIKE 11.
PT DOMAIN 2132 2529 EGF-LIKE 12.
PT DOMAIN 2530 2579 EGF-LIKE 13.
PT DOMAIN 2580 2629 EGF-LIKE 14.
PT DOMAIN 2630 2679 EGF-LIKE 15.
PT DOMAIN 2680 2729 EGF-LIKE 16.
PT DOMAIN 2730 2779 EGF-LIKE 17.
PT DOMAIN 2780 2829 EGF-LIKE 18.
PT DOMAIN 2830 2879 EGF-LIKE 19.
PT DOMAIN 2880 2929 EGF-LIKE 20.
PT DOMAIN 2930 2979 EGF-LIKE 21.
PT DOMAIN 2980 3029 EGF-LIKE 22.
PT DOMAIN 3030 3079 EGF-LIKE 23.
PT DOMAIN 3080 3129 EGF-LIKE 24.
PT DOMAIN 3130 3179 EGF-LIKE 25.
PT DOMAIN 3180 3229 EGF-LIKE 26.
PT DOMAIN 3230 3279 EGF-LIKE 27.
PT DOMAIN 3280 3329 EGF-LIKE 28.
PT DOMAIN 3330 3379 EGF-LIKE 29.
PT DOMAIN 3380 3429 EGF-LIKE 30.
PT DOMAIN 3430 3479 EGF-LIKE 31.
PT DOMAIN 3480 3529 EGF-LIKE 32.
PT DOMAIN 3530 3579 EGF-LIKE 33.
PT DOMAIN 3580 3629 EGF-LIKE 34.
PT DOMAIN 3630 3679 EGF-LIKE 35.
PT DOMAIN 3680 3729 EGF-LIKE 36.
PT DOMAIN 3730 3779 EGF-LIKE 37.
PT DOMAIN 3780 3829 EGF-LIKE 38.
PT DOMAIN 3830 3879 EGF-LIKE 39.
PT DOMAIN 3880 3929 EGF-LIKE 40.
PT DOMAIN 3930 3979 EGF-LIKE 41.
PT DOMAIN 3980 4029 EGF-LIKE 42.
PT DOMAIN 4030 4079 EGF-LIKE 43.
PT DOMAIN 4080 4129 EGF-LIKE 44.
PT DOMAIN 4130 4179 EGF-LIKE 45.
PT DOMAIN 4180 4229 EGF-LIKE 46.
PT DOMAIN 4230 4279 EGF-LIKE 47.
PT DOMAIN 4280 4329 EGF-LIKE 48.
PT DOMAIN 4330 4379 EGF-LIKE 49.
PT DOMAIN 4380 4429 EGF-LIKE 50.
PT DOMAIN 4430 4479 EGF-LIKE 51.
PT DOMAIN 4480 4529 EGF-LIKE 52.
PT DOMAIN 4530 4579 EGF-LIKE 53.
PT DOMAIN 4580 4629 EGF-LIKE 54.
PT DOMAIN 4630 4679 EGF-LIKE 55.
PT DOMAIN 4680 4729 EGF-LIKE 56.
PT DOMAIN 4730 4779 EGF-LIKE 57.
PT DOMAIN 4780 4829 EGF-LIKE 58.
PT DOMAIN 4830 4879 EGF-LIKE 59.
PT DOMAIN 4880 4929 EGF-LIKE 60.
PT DOMAIN 4930 4979 EGF-LIKE 61.
PT DOMAIN 4980 5029 EGF-LIKE 62.
PT DOMAIN 5030 5079 EGF-LIKE 63.
PT DOMAIN 5080 5129 EGF-LIKE 64.
PT DOMAIN 5130 5179 EGF-LIKE 65.
PT DOMAIN 5180 5229 EGF-LIKE 66.
PT DOMAIN 5230 5279 EGF-LIKE 67.
PT DOMAIN 5280 5329 EGF-LIKE 68.
PT DOMAIN 5330 5379 EGF-LIKE 69.
PT DOMAIN 5380 5429 EGF-LIKE 70.
PT DOMAIN 5430 5479 EGF-LIKE 71.
PT DOMAIN 5480 5529 EGF-LIKE 72.
PT DOMAIN 5530 5579 EGF-LIKE 73.
PT DOMAIN 5580 5629 EGF-LIKE 74.
PT DOMAIN 5630 5679 EGF-LIKE 75.
PT DOMAIN 5680 5729 EGF-LIKE 76.
PT DOMAIN 5730 5779 EGF-LIKE 77.
PT DOMAIN 5780 5829 EGF-LIKE 78.
PT DOMAIN 5830 5879 EGF-LIKE 79.
PT DOMAIN 5880 5929 EGF-LIKE 80.
PT DOMAIN 5930 5979 EGF-LIKE 81.
PT DOMAIN 5980 6029 EGF-LIKE 82.
PT DOMAIN 6030 6079 EGF-LIKE 83.
PT DOMAIN 6080 6129 EGF-LIKE 84.
PT DOMAIN 6130 6179 EGF-LIKE 85.
PT DOMAIN 6180 6229 EGF-LIKE 86.
PT DOMAIN 6230 6279 EGF-LIKE 87.
PT DOMAIN 6280 6329 EGF-LIKE 88.
PT DOMAIN 6330 6379 EGF-LIKE 89.
PT DOMAIN 6380 6429 EGF-LIKE 90.
PT DOMAIN 6430 6479 EGF-LIKE 91.
PT DOMAIN 6480 6529 EGF-LIKE 92.
PT DOMAIN 6530 6579 EGF-LIKE 93.
PT DOMAIN 6580 6629 EGF-LIKE 94.
PT DOMAIN 6630 6679 EGF-LIKE 95.
PT DOMAIN 6680 6729 EGF-LIKE 96.
PT DOMAIN 6730 6779 EGF-LIKE 97.
PT DOMAIN 6780 6829 EGF-LIKE 98.
PT DOMAIN 6830 6879 EGF-LIKE 99.
PT DOMAIN 6880 6929 EGF-LIKE 100.
PT DOMAIN 6930 6979 EGF-LIKE 101.
PT DOMAIN 6980 7029 EGF-LIKE 102.
PT DOMAIN 7030 7079 EGF-LIKE 103.
PT DOMAIN 7080 7129 EGF-LIKE 104.
PT DOMAIN 7130 7179 EGF-LIKE 105.
PT DOMAIN 7180 7229 EGF-LIKE 106.
PT DOMAIN 7230 7279 EGF-LIKE 107.
PT DOMAIN 7280 7329 EGF-LIKE 108.
PT DOMAIN 7330 7379 EGF-LIKE 109.
PT DOMAIN 7380 7429 EGF-LIKE 110.
PT DOMAIN 7430 7479 EGF-LIKE 111.
PT DOMAIN 7480 7529 EGF-LIKE 112.
PT DOMAIN 7530 7579 EGF-LIKE 113.
PT DOMAIN 7580 7629 EGF-LIKE 114.
PT DOMAIN 7630 7679 EGF-LIKE 115.
PT DOMAIN 7680 7729 EGF-LIKE 116.
PT DOMAIN 7730 7779 EGF-LIKE 117.
PT DOMAIN 7780 7829 EGF-LIKE 118.
PT DOMAIN 7830 7879 EGF-LIKE 119.
PT DOMAIN 7880 7929 EGF-LIKE 120.
PT DOMAIN 7930 7979 EGF-LIKE 121.
PT DOMAIN 7980 8029 EGF-LIKE 122.
PT DOMAIN 8030 8079 EGF-LIKE 123.
PT DOMAIN 8080 8129 EGF-LIKE 124.
PT DOMAIN 8130 8179 EGF-LIKE 125.
PT DOMAIN 8180 8229 EGF-LIKE 126.
PT DOMAIN 8230 8279 EGF-LIKE 127.
PT DOMAIN 8280 8329 EGF-LIKE 128.
PT DOMAIN 8330 8379 EGF-LIKE 129.
PT DOMAIN 8380 8429 EGF-LIKE 130.
PT DOMAIN 8430 8479 EGF-LIKE 131.
PT DOMAIN 8480 8529 EGF-LIKE 132.
PT DOMAIN 8530 8579 EGF-LIKE 133.
PT DOMAIN 8580 8629 EGF-LIKE 134.
PT DOMAIN 8630 8679 EGF-LIKE 135.
PT DOMAIN 8680 8729 EGF-LIKE 136.
PT DOMAIN 8730 8779 EGF-LIKE 137.
PT DOMAIN 8780 8829 EGF-LIKE 138.
PT DOMAIN 8830 8879 EGF-LIKE 139.
PT DOMAIN 8880 8929 EGF-LIKE 140.
PT DOMAIN 8930 8979 EGF-LIKE 141.
PT DOMAIN 8980 9029 EGF-LIKE 142.
PT DOMAIN 9030 9079 EGF-LIKE 143.
PT DOMAIN 9080 9129 EGF-LIKE 144.
PT DOMAIN 9130 9179 EGF-LIKE 145.
PT DOMAIN 9180 9229 EGF-LIKE 146.
PT DOMAIN 9230 9279 EGF-LIKE 147.
PT DOMAIN 9280 9329 EGF-LIKE 148.
PT DOMAIN 9330 9379 EGF-LIKE 149.
PT DOMAIN 9380 9429 EGF-LIKE 150.
PT DOMAIN 9430 9479 EGF-LIKE 151.
PT DOMAIN 9480 9529 EGF-LIKE 152.
PT DOMAIN 9530 9579 EGF-LIKE 153.
PT DOMAIN 9580 9629 EGF-LIKE 154.
PT DOMAIN 9630 9679 EGF-LIKE 155.
PT DOMAIN 9680 9729 EGF-LIKE 156.
PT DOMAIN 9730 9779 EGF-LIKE 157.
PT DOMAIN 9780 9829 EGF-LIKE 158.
PT DOMAIN 9830 9879 EGF-LIKE 159.
PT DOMAIN 9880 9929 EGF-LIKE 160.
PT DOMAIN 9930 9979 EGF-LIKE 161.
PT DOMAIN 10000 10000 EGF-LIKE 162.

```

```

Query Match 0.64; Score 18; DB 1; Length 3312;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2495 GLAVGLDPGYNPFCM 2512

```

```

DB 2658 GLAVGLDPGYNPFCM 2675
RESULT 8
ID CLRS RAT STANDARD; PRT; 3313 AA.
AC 088278;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF-like seven-pass G-type receptor 3 precursor (Multiple
DE epidermal growth factor-like domains 2).
GN CELSR3 OR MEKFP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC cerebellum, olfactory bulb, cerebral cortex, hippocampus and
CC brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB011528; BAA32459.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR000152; Abx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cye_rich.
DR Pfam; PF000002; 7tm_2; 1.
DR Pfam; PF000028; cadherin_9.
DR Pfam; PF00008; EGF_6.
DR Pfam; PF01825; GPS_1.
DR Pfam; PF02793; HRM_1.
DR Pfam; PF00054; laminin_G_1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR00243; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; HormR; 1.

```





AC 095450; (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 42, Last annotation update)  
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase  
 with thrombospondin motifs 2) (ADAM-TS2) (Procollagen I/II  
 amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2 OR PCINP OR PCPIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.  
 RC TISSUE=Skin;  
 RX MEDLINE=99347935; PubMed=10417273;  
 RA Collage A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,  
 RA Wierzelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,  
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;  
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis  
 are caused by mutations in the procollagen I N-proteinase gene.";  
 RL Am. J. Hum. Genet. 65:308-317(1999).  
 CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
 to fibril assembly. Does not act on type III collagen. May also  
 play a role in development that is independent of its role in  
 collagen biosynthesis.  
 CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
 at Ala-|-Gln.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
 COLLAGEN TYPE XIV (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 matrix (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=LPNPI;  
 CC IsoId=095450-1; Sequence=Displayed;  
 CC Name=SPNPI;  
 CC IsoId=095450-2; Sequence=VSP\_005497, VSP\_005498;  
 CC Note=Has no significant N-procollagen peptidase activity;  
 CC TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON  
 AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.  
 CC -1- DOMAIN: THE SPACER DOMAIN WITH THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE (By  
 SIMILARITY).  
 CC -1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos  
 syndrome type VIIC (EDS-VIIC), a recessively inherited connective-  
 tissue disorder characterized clinically by severe skin fragility  
 and joint hypermobility and biochemically by the presence in skin  
 of procollagen incompletely processed at the N-terminus.  
 CC -1- SIMILARITY: Belongs to peptidase family M12B.  
 CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 4 TSP type-1 domains.  
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AJ003125; CA05880.1; -.  
 DR MEROPS: M12.301; -.  
 DR Genew; HGNC:218; ADAMTS2.  
 DR MIM; 604539; -.  
 DR MIM; 225410; -.  
 DR GO; GO:0008133; F:collagenase activity; TAS.

DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR002870; Pep\_M12B\_propep.  
 DR InterPro; IPR001590; Reprlysin.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR006025; Zn\_M12B\_peptide.  
 DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam; PF01421; Reprlysin; 1.  
 DR SMART; SM00209; TSP1; 4.  
 DR PROSITE; PS50215; ADAM\_MEROPS; 1.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE; PS50214; DISINTEGRIN\_2; FALSE\_NEG.  
 DR PROSITE; PS50092; TSP1; 4.  
 DR PROSITE; PS00427; ZINC\_PROTEASE; FALSE\_NEG.  
 DR Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;  
 KW Alternative splicing; Ehlers-Danlos syndrome.  
 KW SIGNAL  
 FT SIGNAL 1 29  
 FT PROPEP 30 253  
 FT CHAIN 254 1211  
 FT DOMAIN 254 479  
 FT DOMAIN 480 560  
 FT DOMAIN 561 616  
 FT DOMAIN 618 722  
 FT SITE 651 693  
 FT SITE 723 851  
 FT DOMAIN 854 912  
 FT DOMAIN 914 971  
 FT DOMAIN 975 1029  
 FT DOMAIN 1063 1090  
 FT DOMAIN 40 43  
 FT DOMAIN 185 188  
 FT METAL 408 408  
 FT ACT SITE 409 409  
 FT METAL 412 412  
 FT METAL 418 418  
 FT CARBOHYD 112 112  
 FT CARBOHYD 251 251  
 FT CARBOHYD 949 949  
 FT CARBOHYD 993 993  
 FT CARBOHYD 1031 1031  
 FT CARBOHYD 1098 1098  
 FT CARBOHYD 1145 1145  
 FT CARBOHYD 1150 1150  
 FT VASPLIC 544 566  
 FT FT  
 FT VASPLIC 567 1211  
 FT FT  
 FT SEQUENCE 1211 AA; 134722 MW; BECEPF25C23CAD2D CRC64;  
 SQ  
 Query Match 0.4%; Score 13; DB 1; Length 1211;  
 Best Local Similarity 100.0%; Pred. NO. 0.089;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 LLLLLLLLLPPL 30  
 DB 16 LLLLLLLLLPPL 28  
 RESULT 12  
 CD87 DROME STANDARD; FRT; 1967 AA.  
 AC Q9VG65; Q9VG65; (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 42, Last annotation update)  
 DE 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE Cadherin 87A precursor.  
 GN CAD87A OR CG6977.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC EMBL; AE003693; AAF54717.3; -  
 CC EMBL; AY128505; AAM75098.1; -  
 CC HSSP; P5116; INCL  
 CC Flybase: FBgn0037963; Cad87A.  
 CC GO; GO:0016021; C: integral to membrane; NAS.  
 CC GO; GO:0008014; F: calcium-dependent cell adhesion molecule ac. .; NAS.  
 CC GO; GO:0016339; P: calcium-dependent cell-cell adhesion; NAS.  
 CC InterPro; IPR002126; Cadherin.  
 CC Pfam; PF00028; cadherin; 2.  
 CC PRINTS; PR00205; CADHERIN.  
 CC SMART; SM00112; CA; 14.  
 CC PROSITE; PS00232; CADHERIN 1; 3.  
 CC PROSITE; PS0268; CADHERIN 2; 14.  
 CC Hypothetical protein; Cell adhesion; Glycoprotein; Transmembrane;  
 CC Calcium; Calcium-binding; Repeat; Signal; Multigene family.  
 CC SIGNAL  
 CC CHAIN  
 CC ? 1967  
 CC ? 1767  
 CC TRANSMEM  
 CC 1768 1788  
 CC 1789 1967  
 CC 20 124  
 CC 125 237  
 CC 238 350  
 CC 351 464  
 CC 465 661  
 CC 662 766  
 CC 767 870  
 CC 871 990  
 CC 991 1095  
 CC 1096 1203  
 CC 1204 1310  
 CC 1311 1423  
 CC 1424 1545  
 CC 1546 1669  
 CC 31 31  
 CC 69 69  
 CC 195 195  
 CC 416 416  
 CC 722 722  
 CC 753 753  
 CC 1031 1031  
 CC 1041 1041  
 CC 1103 1103  
 CC 1155 1155  
 CC 1209 1209  
 CC 1317 1317  
 CC 1341 1341  
 CC 1484 1484  
 CC 1568 1568  
 CC 1683 1683  
 CC 763 763  
 CC 1102 1102  
 CC 1642 1642  
 CC 1682 1682  
 CC 1967 AA; 216689 MW; 7164858580CP41 CRC64;  
 CC  
 CC Query Match 0.4%; Score 13; DB 1; Length 1967;  
 CC Best Local Similarity 100.0%; Pred. No. 0.13;  
 CC Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 910 TVLDVNDNPVPR 922  
 CC |||||  
 CC Db 1534 TVLDVNDNPVPR 1546

RT NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Manamickandes P.G., Scherer S.B., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Champs P., Pfeiffer B.D.,  
 RA Wen K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Bouch M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Canter A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosløk A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 RA Jalali M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,  
 RA Palazolo M., Piatan G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sider-Kramos I., Simpson M., Skupski M.P., Smith T.,  
 RA Sjöter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Waastman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=Berkley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Mathews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.B.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [3]  
 RP SEQUENCE OF 435-1967 FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Paclet J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.  
 CC They preferentially interact with themselves in a homophilic  
 CC manner in connecting cells (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
 CC -1- SIMILARITY: Contains 14 cadherin domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



```

RLA0_METKA
ID METKA STANDARD; PRT; 357 AA.
AC 08TX50;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acidic ribosomal protein P0 homolog (L10B).
GN RPLP0 OR MK0826.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxId=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Sleasman A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natchale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Steetler K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
CC -1- FUNCTION: Ribosomal protein P0 is the functional equivalent of
CC E. coli protein L10.
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; AE010373; AM02039.1; ALT_INIT.
DR HAMAP; MF_00280; -1.
DR InterPro; IPR001790; Ribosomal_L10.
DR Pfam; PF00466; Ribosomal_L10; 1.
DR K01001; Ribosomal protein, Complete proteome.
SQ SEQUENCE 357 AA; 39250 MW; 470294320ADBBESC CRC64;

Query Match 0.4%; Score 12; DB 1; Length 357;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 EEEEEEEEA 2754
DB 338 EEEEEEEEA 349

RESULT 14
PTPX MACNE STANDARD; PRT; 1013 AA.
AC 002655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein-tyrosine phosphatase X precursor (EC 3.1.3.48) (R-PTP-X)
DE (M1851).
GN PTPRX2.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxId=9545;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RX MEDLINE=97254813; PubMed=9100223;
RA Lagase J., Jellinek L., Sexson S., Lofton-Day C.B., Breining J.,
RA Sheppard P., Kindsvogel W., Hagopian W.A.;
RT "An islet-cell protein tyrosine phosphatase is a likely precursor to

```

```

RT the 37-kDa autoantigen in type 1 diabetes: human and macaque
RT sequences, tissue distribution, unique and shared epitopes, and
RT predictive autoantibodies."
RL Mol. Med. 3:163-173 (1997).
CC -1- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND
CC PANCREATIC ENDOCRINE CELLS.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: BRAIN, PROSTATE, PANCREATIC ISLETS. LOWER
CC EXPRESSION IN SPINAL CORD, THYROID, ADRENAL MEDULLA AND
CC GASTROINTESTINAL TRACT. MULTIPLE PROTEOLYTIC CLEAVAGE AT
CC -1- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT
CC CONSECUTIVE BASIC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; U91574; AAC51186.1; -.
DR HSSP; P18052; TYRO.
DR InterPro; IPR000387; TYR phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTPHPTASR.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
DR K01001; Hydroxylase; Receptor; Glycoprotein; Signal; Transmembrane.
FT SIGNAL 1 19
FT CHAIN 20 1013
FT DOMAIN 20 613
FT TRANSMEM 614 634
FT DOMAIN 635 1013
FT ACT SITE 943 943
FT SITE 425 426
FT CARBOHYD 562 562
SQ SEQUENCE 1013 AA; 11190 MW; 4808D43937A2EF59 CRC64;

Query Match 0.4%; Score 12; DB 1; Length 1013;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PLLLLLLLLLP 27
DB 4 PLLLLLLLLLP 15

RESULT 15
PTPX HUMAN STANDARD; PRT; 1015 AA.
AC 092932; Q92662;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-tyrosine phosphatase X precursor (EC 3.1.3.48) (R-PTP-X)
DE (Islet cell autoantigen related protein) (ICAPAR) (IAR) (Phogfin).
GN PTPRX2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=97032784; PubMed=8878534;

```

RA Kawasaki E., Hutton J.C., Eisenbarth G.S.;  
 RT "Molecular cloning and characterization of the human transmembrane  
 RT protein tyrosine phosphatase homologue, plogrin, an autoantigen of  
 RT type 1 diabetes.";  
 RL Biochem. Biophys. Res. Commun. 227:440-447(1996).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97127415; PubMed=8954911;  
 RA Smith P.D., Barker K.T., Wang J., Lu Y.-J., Shipley J., Crompton M.R.;  
 RT "ICAR, a novel member of a new family of transmembrane, tyrosine  
 RT phosphatase-like proteins.";  
 RL Biochem. Biophys. Res. Commun. 229:402-411(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Brain, and Pancreas;  
 RX MEDLINE=96394649; PubMed=8798755;  
 RA Cui L., Yu W.-P., de Aizpuru H.J., Schmidt R.S., Pallen C.J.;  
 RT "Cloning and characterization of islet cell antigen-related protein-  
 RT tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen  
 RT in insulin-dependent diabetes.";  
 RL J. Biol. Chem. 271:24817-24823(1996).  
 CC -1- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND  
 CC PANCREATIC ENDOCRINE CELLS. OPTIMUM ACTIVITY IS MEASURED AT PH  
 CC 4.5.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND PANCREAS. LOWER  
 CC LEVELS IN TRACHEA, PROSTATE, STOMACH AND SPINAL CHORD.  
 CC -1- DOMAIN: THE CYTOPLASMIC DOMAIN APPEARS TO CONTRAIN THE  
 CC AUTOANTIGENIC EPITOPES.  
 CC -1- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT  
 CC CONSECUTIVE BASIC RESIDUES.  
 CC -1- DISEASE: AUTOANTIGEN IN INSULIN-DEPENDENT DIABETES MELLITUS  
 CC (IDDM).  
 CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U66702; AAC50742.1; -;  
 DR EMBL: Y08569; CA69880.1; -;  
 DR EMBL: AF007555; AAB63600.1; -;  
 DR PIR: JC5062; JC5062.  
 DR PIR: JC5263; JC5263.  
 DR HSP: P18052; IYFO.  
 DR Genew: HGNC:9677; PTPRN2.  
 DR MIM: 601698; -;  
 DR GO: GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO: GO:0005001; F:transmembrane receptor protein tyrosine pho. . .; TAS.  
 DR GO: GO:006470; P:protein amino acid dephosphorylation; TAS.  
 DR InterPro: IPR000387; TYR phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase; 1.  
 DR PRINTS: PR00700; PRTYPHRTASE.  
 DR SMART: SM00194; PTPC; 1.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR HydroLase; Receptor; Glycoprotein; Signal; Transmembrane.  
 KW HydroLase; Receptor; Glycoprotein; Signal; Transmembrane.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1015 PROTEIN-TYROSINE PHOSPHATASE X.  
 FT DOMAIN 22 615 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 616 636 POTENTIAL.  
 FT DOMAIN 637 1015 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 770 1004 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 945 945

FT SITE 427 428  
 FT CARBOXYD 564 564  
 FT MUTAGEN 945 945  
 FT CONFLICT 208 208  
 FT CONFLICT 247 247  
 FT CONFLICT 323 323  
 FT CONFLICT 325 325  
 SQ SEQUENCE 1015 AA; 111281 MW; 82AC3CC9F59AC97 CRC64;  
 Query Match 0.4%; Score 12; DB 1; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 16 P|||||L|L|L|P 27  
 Db 6 P|||||L|L|L|P 17  
 Search completed: February 11, 2004, 16:06:43  
 Job time : 28 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:03:00 ; Search time 75 Seconds  
(without alignments)  
10057.170 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923  
Sequence: 1 MESPATGVLPPTPPPLLL.....AGTVDSSGSRFLPFLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phase:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacterioplasmid:\*  
17: SP\_archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	1.1	32	11	Q9J1U2
2	24	0.8	25	11	Q9J1U4
3	19	0.7	34	6	O45539
4	18	0.6	138	13	Q9YQ08
5	18	0.6	1766	13	O8AM45
6	15	0.5	138	13	Q9YQ09
7	13	0.4	311	11	Q8VE43
8	13	0.4	382	11	Q9D0M2
9	13	0.4	382	11	O8C1A0
10	13	0.4	414	6	O8HZR0
11	13	0.4	436	11	O8BMR6
12	13	0.4	467	4	Q9BTY2
13	13	0.4	467	4	O8BNK4
14	13	0.4	467	4	Q9JUM5
15	13	0.4	633	6	O8HZR1
16	12	0.4	176	11	Q8QZT0

17	12	0.4	176	11	Q8C1M2	Q8C1M2 mus musculus
18	12	0.4	239	6	Q9XSK7	Q9XSK7 bos taurus
19	12	0.4	271	11	Q8CDY2	Q8CDY2 mus musculus
20	12	0.4	293	10	Q9R1U5	Q9R1U5 arabidopsis
21	12	0.4	315	4	Q9EHG7	Q9EHG7 homo sapien
22	12	0.4	330	11	Q8C1A4	Q8C1A4 mus musculus
23	12	0.4	334	11	Q8CIU8	Q8CIU8 mus musculus
24	12	0.4	401	11	Q9CWC8	Q9CWC8 mus musculus
25	12	0.4	401	11	Q8CIU9	Q8CIU9 mus musculus
26	12	0.4	401	11	O8BSN5	O8BSN5 mus musculus
27	12	0.4	403	4	Q9BUL5	Q9BUL5 homo sapien
28	12	0.4	538	4	Q9NMQ7	Q9NMQ7 homo sapien
29	12	0.4	540	4	O8N0Z9	O8N0Z9 homo sapien
30	12	0.4	608	6	O97554	O97554 oryctolagus
31	12	0.4	608	11	Q9SU12	Q9SU12 arabidopsis
32	12	0.4	629	11	Q9RT51	Q9RT51 mus musculus
33	12	0.4	793	11	O70246	O70246 mus musculus
34	12	0.4	813	11	Q8BQC3	Q8BQC3 mus musculus
35	12	0.4	896	11	Q9R093	Q9R093 ratius norv
36	12	0.4	900	17	O8PUB5	O8PUB5 methanosarc
37	12	0.4	986	4	O9Y419	O9Y419 homo sapien
38	12	0.4	986	4	O8N4T5	O8N4T5 homo sapien
39	12	0.4	998	4	O9Y416	O9Y416 homo sapien
40	12	0.4	1032	11	Q9Z1T4	Q9Z1T4 ratius norv
41	12	0.4	1270	4	O96JH2	O96JH2 homo sapien
42	12	0.4	1378	11	O35850	O35850 ratius norv
43	12	0.4	4169	4	O8CTU4	O8CTU4 homo sapien
44	11	0.4	59	11	Q9R1N5	Q9R1N5 mus musculus
45	11	0.4	75	10	Q9STQ5	Q9STQ5 arabidopsis

## ALIGNMENTS

## RESULT 1

ID Q9J1U2 PRELIMINARY; PRT; 32 AA.  
AC Q9J1U2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 22, Last annotation update)  
DE Flamingol (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fischer344; TISSUE=Testis;  
RX MEDLINE=20114371; PubMed=10650949;  
RA Johnson K.J., Patel S.R., Boekelheide K.;  
RT "Multiple cadherin superfamily members with unique expression profiles are produced in rat testis";  
RL Endocrinology 141:675-683(2000).  
DR EMBL; AF177697; AAF87072.1; -.  
FT NON TER 1 32  
FT MATCH 32  
SQ SEQUENCE 32 AA; 3554 MW; 60ABB46A7B62616B CRC64;

Query Match 777 DOVSYTLATFARDNGIPKSDTTTLEILVNDV 808  
Best Local Similarity 100.0%; Pred. No. 3, 1e-23;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 DOVSYTLATFARDNGIPKSDTTTLEILVNDV 32

## RESULT 2

ID Q9J1U4 PRELIMINARY; PRT; 25 AA.  
AC Q9J1U4;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Flamingo1 (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer344; TISSUE=Testis;  
 RX MEDLINE=20114371; PubMed=10650949;  
 RA Johnson K.J., Patel S.R., Boekelheide K.;  
 RT "Multiple cadherin superfamily members with unique expression profiles  
 are produced in rat testis."  
 RL Endocrinology 141:675-683(2000).  
 DR EMBL; AF177695; AAF87070.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 25 25  
 SQ SEQUENCE 25 AA; 2605 MW; 09940DE0F9P95794 CRC64;

Query Match 0.8%; Score 24; DB 11; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-15;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 RAODGGRPLSNVSGLVTVQVLDI 498  
 DB 2 RAODGGRPLSNVSGLVTVQVLDI 25

## RESULT 3

046539 PRELIMINARY; PRT; 34 AA.  
 AC 046539;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Cadherin-like protein (Fragment).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 NCBI\_TaxID=9823;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RA Lutz D.A., Zheng J.J.;  
 RT "Expression of multiple cadherins in adult retinal pigment epithelial  
 (RPE) cells."  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF037464; AAB92582.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 34 34  
 SQ SEQUENCE 34 AA; 3648 MW; 5FCF1719B718F855 CRC64;

Query Match 0.7%; Score 19; DB 6; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 YTLAITARDNGIPKSDPT 799  
 DB 5 YTLAITARDNGIPKSDPT 23

## RESULT 4

09YGO8 PRELIMINARY; PRT; 138 AA.  
 AC 09YGO8;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Cadherin (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.  
 NCBI\_TaxID=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Miskavich F., Zhu Y., Ranscht B., Sanes J.R.;  
 RT "Expression of Multiple Cadherins and Catenins in the Chick Optic  
 Tectum."  
 RL Mol. Cell. Neurosci. 0:0-0(1999).  
 DR EMBL; AF098479; AAD16269.1; -.  
 DR HSSP; P09803; 1EDH.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 1.  
 DR SMART; SM00112; Ca; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 1.  
 DR PROSITE; PS50268; CADHERIN\_2; 2.  
 KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 138 138  
 SQ SEQUENCE 138 AA; 14523 MW; E03B8D50074F48E6 CRC64;

Query Match 0.6%; Score 18; DB 13; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 NDNAPIFVSTPPOATYLE 516  
 DB 34 NDNAPIFVSTPPOATYLE 51

## RESULT 5

08AW45 PRELIMINARY; PRT; 1766 AA.  
 AC 08AW45;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Sl:b6L08.1 (Novel protein similar to cadherin, EGF LAG seven-pass  
 G-type receptor (CELSR)) (Fragment).  
 GN Sl:B6L08.1.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_TaxID=7955;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Babbage A.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL645790; CAD58738.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 1766 1766  
 SQ SEQUENCE 1766 AA; 197485 MW; 80CANA18F5A497F CRC64;

Query Match 0.6%; Score 18; DB 13; Length 1766;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 VLPPDNDICLRPCENYM 1241  
 DB 117 VLPPDNDICLRPCENYM 134

## RESULT 6

09YGO9 PRELIMINARY; PRT; 138 AA.  
 AC 09YGO9;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Cadherin (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Miskelich F., Zhu Y., Ranscht B., Sanes J.R.;
RT "Expression of Multiple Cadherins and Catenins in the Chick Optic
RL Tectum."
DR EMBL: AF098478; AAD1268.1; -.
DR HSRP; P15116; INCI.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 1.
DR PROSITE; PS00232; CADHERIN_1; 1.
DR PROSITE; PS00268; CADHERIN_2; 2.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15282 MW; A68C982E27CB290 CRC64;

Query Match 0.4%; Score 15; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 PPFTSVLOISATRPD 843
DB 53 PPFTSVLOISATRPD 67

RESULT 7
O8VE43 PRELIMINARY; PRT; 311 AA.
AC O8VE43;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Hypothetical protein.
GN BC019776.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RL EMBL; BC019776; AAH19776.1; -.
DR EMBL; BC024445; AAH24445.1; -.
DR EMBL; BC026646; AAH26646.1; -.
DR EMBL; AK034855; BAC28855.1; -.
DR MGD; MGI:2384806; BC019776.
KW Hypothetical protein.
SQ SEQUENCE 311 AA; 34529 MW; C806268E2B6DFA0 CRC64;

Query Match 0.4%; Score 13; DB 11; Length 311;

```

```

Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 PPPPLLLLLL 25
DB 28 PPPPLLLLLL 40

RESULT 8
O9D0M2 PRELIMINARY; PRT; 382 AA.
ID O9D0M2;
AC O9D0M2;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE 2310021G01RIK protein (Similar to C-MYC target JF01).
GN 2310021G01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Stebbins F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazarelli J., Mombasarte P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RL EMBL; AK011289; BAB27519.1; -.
DR EMBL; AK075882; BAC36027.1; -.
DR MGD; MGI:1914203; 2310021G01RIK.
SQ SEQUENCE 382 AA; 43837 MW; 5192063588CD40 CRC64;

Query Match 0.4%; Score 13; DB 11; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2740 SDSEEEEEEEEE 2752
DB 103 SDSEEEEEEEEE 115

RESULT 9
O8C1A0 PRELIMINARY; PRT; 382 AA.
ID O8C1A0;
AC O8C1A0;
DT 01-MAR-2003 (TRENBLREL. 23, Created)

```

DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
 DE Similar to C-MYC target JPOI.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL: AK028671; BAC26058.1; -  
 SQ SEQUENCE 382 AA; 43836 MW; 33F004B500ED9BE CRC64;

Query Match 0.4%; Score 13; DB 11; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2740 SDSEEEEEEEER 2752  
 DB 103 SDSEEEEEEEER 115

## RESULT 10

Q8H2R0 ID Q8H2R0 PRELIMINARY; PRT; 414 AA.  
 AC Q8H2R0;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DE Cyclooxxygenase-1 variant.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 NCBI\_TaxID=9615;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12242329;  
 RA Chandrasekharan N.V., Dai H., Roos K.L.T., Evanson N.K., Tomsak J.,  
 RA Elton T.S., Simmons D.L.;  
 RT "COX-3, a cyclooxxygenase-1 variant inhibited by acetaminophen and  
 RT other analgesic/antipyretic drugs: Cloning, structure, and  
 RT expression."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13926-13931(2002).  
 DR EMBL: AF535139; AAN38739.1; -  
 SQ SEQUENCE 414 AA; 47510 MW; 75B160CE7390086E CRC64;

Query Match 0.4%; Score 13; DB 6; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 0.0025;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PLLLLLLLLPPP 28  
 DB 41 PLLLLLLLLPPP 53

## RESULT 11

Q8BNR6 ID Q8BNR6 PRELIMINARY; PRT; 436 AA.  
 AC Q8BNR6;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DE Hypothetical protein.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL: AK080819; BAC38034.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 436 AA; 47342 MW; 18D92C1B2812A6F CRC64;

Query Match 0.4%; Score 13; DB 11; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 0.0026;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PTPPPPLLLLLL 23  
 DB 29 PTPPPPLLLLLL 41

## RESULT 12

Q9BTY2 ID Q9BTY2 PRELIMINARY; PRT; 467 AA.  
 AC Q9BTY2;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
 DE Similar to fucosidase, alpha-L-1, tissue.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC003060; AAH03060.1; -  
 DR InterPro: IPR000933; Alpha\_L\_fucos.  
 DR Pfam: PF01120; Alpha\_L\_fucos. 1.  
 DR PRINTS: PR00741; GLHYDRLASE29.  
 DR PROSITE: PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 SQ SEQUENCE 467 AA; 54060 MW; FF3B0DA9D818AD85 CRC64;

Query Match 0.4%; Score 13; DB 4; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PLLLLLLLLPPP 28  
 DB 12 PLLLLLLLLPPP 24

## RESULT 13

Q8NBK4 ID Q8NBK4 PRELIMINARY; PRT; 467 AA.  
 AC Q8NBK4;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
 DE Hypothetical protein PLACE1007878.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
 RA Nagahari K., Sugano S., Isogai T.;

RT "HRI human cDNA sequencing project";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A075458; BAC11633.1; -  
 DR InterPro; IPR000933; Alpha\_L\_fucos.  
 DR Pfam; PF01120; Alpha\_L\_fucos; 1.  
 DR PRINTS; PR00741; GLHYDRLASE29.  
 DR PROSITE; PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 KW Hypothetical protein: 54044 MW; 933BDB9858BDD5 CRC64;  
 SQ SEQUENCE 467 AA; 54044 MW; 933BDB9858BDD5 CRC64;  
 Query Match 0.4%; Score 13; DB 4; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 12 PLLLLLLLLPPP 24

RESULT 14  
 Q9UUM5 PRELIMINARY; PRT; 467 AA.  
 AC Q9UUM5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE DJ20N2.5 (Novel protein similar to fucosidase, alpha-L-1, tissue  
 (EC 3.2.1.51, alpha-L-fucosidase fucohydrolase)).  
 GN DJ20N2.5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kay M.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031320; CMB53746.1; -  
 DR InterPro; IPR000933; Alpha\_L\_fucos.  
 DR Pfam; PF01120; Alpha\_L\_fucos; 1.  
 DR PRINTS; PR00741; GLHYDRLASE29.  
 DR PROSITE; PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 SQ SEQUENCE 467 AA; 54066 MW; A42AA6B369A3AC39 CRC64;  
 Query Match 0.4%; Score 13; DB 4; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 12 PLLLLLLLLPPP 24

RESULT 15  
 O8HZR1 PRELIMINARY; PRT; 633 AA.  
 AC O8HZR1;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Cyclooxxygenase.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_Taxid=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pubmed-12242329;  
 RA Chandrasekharan N.V., Dai H., Roos K.L.T., Evanson N.K., Tomask J.,  
 Elton T.S., Simmons D.L.;  
 RT "COX-3, a cyclooxxygenase-1 variant inhibited by acetaminophen and  
 other analgesic/antipyretic drugs: Cloning, structure, and  
 expression.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:13926-13931(2002).  
 DR EMBL; AF535138; AAN33049.1; -  
 SQ SEQUENCE 633 AA; 72502 MW; D338221B976BDB7E CRC64;  
 Query Match 0.4%; Score 13; DB 6; Length 633;  
 Best Local Similarity 100.0%; Pred. No. 0.0036;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 41 PLLLLLLLLPPP 53

Search completed: February 11, 2004, 16:08:09  
 Job time : 76 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 05:45:30 ; Search time 22141 Seconds  
(without alignments)  
5400,788 Million cell updates/sec

Title: US-09-916-849a-3  
Perfect score: 2923  
Sequence: 1 MRSPTGVLPPLPPPLLL.....ACTVDESSGSEFLFFNLH 2923

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:  
MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09916849/runat.11022004.143917.21352/app.query.fasta.1.3079  
-DB=GenEmbl -QFMT=fastap -SUPPLX=011go.rge -MINMATCH=0.1 -LOOPT=0.5 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=011go -TRANS=human0.cdi -LIST=45  
-DOCLIGN=200 -THR.SCOR=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro  
-NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849@cgn.1.1.15745 @runat.11022004.143917.21352 -NCPU=6 -ICPU=3  
-NO.WMAP -LARGEBQUERY -NEG.SCOR=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_scs:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mus:\*  
20: em\_mmu:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_to:\*  
27: em\_tss:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ey:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	8772	6	AX224128 Sequence
2	2923	100.0	10531	6	AX549238 Sequence
3	2923	100.0	10531	9	AF234887 Homo sapi
4	2837	97.1	8871	6	AX224126 Sequence
5	2408	82.4	8924	9	D87469 Human mRNA
6	1958	67.0	9321	6	AX367102 Sequence
7	1103	37.7	24370	6	AX646739 Sequence
8	1103	37.7	24370	9	AB065955 Homo sapi
9	1103	37.7	169241	9	AL390252 Homo DNA
10	1103	37.7	227194	2	BX284647 Homo sapi
11	1091	37.3	116603	2	AL157901 Homo sapi
12	609	20.8	2821	9	AK091437 Homo sapi
13	482	16.5	89129	10	AL672200 Mouse DNA
14	482	16.5	234932	2	AC093365 Mus muscu
15	347	11.9	3912	6	AX098823 Sequence
16	305	10.4	235320	2	AC106175 Rattus no
17	305	10.4	235327	2	AC095463 Rattus no
18	305	10.4	248059	2	AC113756 Rattus no
19	305	10.4	254975	2	AC095125 Rattus no
20	209	7.2	932	9	HSA323766 Homo sapi
21	208	7.1	1007	10	AY121289 Homo sapi
22	187	6.4	6699	10	AB011529 Rattus no
23	168	5.7	8950	10	AB028499 Mus muscu
24	116	4.0	2653	10	AF031573 Mus muscu
25	109	3.7	253371	2	AC143681 Macaca mu
26	92	3.1	285	6	AX524129 Sequence
27	92	3.1	285	6	AX552867 Sequence
28	87	3.0	3343	10	BC005499 Mus muscu
29	81	2.8	271	11	G16013 human SRS C
30	53	1.8	300	11	G37263 SHGC-57403
31	44	1.5	159860	2	AC110425 Rattus no
32	43	1.5	26451	2	AC143184 Macaca mu
33	34	1.2	132775	10	AL671899 Mouse DNA
34	32	1.1	97	10	AF177697 Rattus no
35	29	1.0	167579	5	AL645789 zebrafish
36	24	0.8	77	10	AF177695 Rattus no
37	23	0.8	10195	6	AF0094 Sequence 4
38	23	0.8	10919	10	AF031572 Mus muscu
39	23	0.8	238919	2	AC139513 Mus muscu
40	23	0.8	253297	2	AC134076 Rattus no
41	23	0.8	260331	2	AC135400 Rattus no
42	23	0.8	263730	2	AC094314 Rattus no
43	21	0.7	135	6	AR025222 Sequence
44	21	0.7	135	6	AR070009 Sequence
45	21	0.7	135	6	IS1001 Sequence 15

RESULT 1

## ALIGNMENTS

AX224128  
LOCUS AX224128 8772 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 3 from Patent WO0161003.  
ACCESSION AX224128  
VERSION AX224128.1 GI:15554400  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1  
Testa, T.T.  
The human homologue of flamingo  
Patent: WO 0161003-A 3 23-AUG-2001;  
SMITHKLINE BEECHAM PLC (GB)  
location/Qualifiers  
1. 8772  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

FEATURES  
source  
1. 8772  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 1696 a 2814 c 2564 g 1698 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 8772  
Score: 2923.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-916-849a-3 (1-2923) x AX224128 (1-8772)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
Db 1 ATGGAGAGCCCGGCGACCGGCTCCCTCCCAAGCGCGCGCGCTGCTGCTG 60  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
Db 61 TTGCTGCTGCTGCTGCGCGCGCGCTATTGGAGACCAAGTGGGGCCCTGCTGCTG 120  
QY 41 G1SerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTTPLeuCysProSer 60  
Db 121 GGGTCCAGGGAGACAGGCTCTTCGGGGGCTGCGCGCCCAAGGGCTGCTGCTGCAATCC 180  
QY 61 Ser1aSerAsnLeuTTPLeuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80  
Db 181 TCGAGCTCGAAGCTCTGGCTTACACAGCGCTGAGGGATGGGGCACTGAGCTGACT 240  
QY 81 G1YH1sLeuValProH1sH1sAspGlyLeuArgValTTPCysProGlnSerGlnAlaHis 100  
Db 241 GGGCACTGGTATCCCAAGGAGCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300  
QY 101 I1eProLeuProProAlaProGlnGlyCysArgProThrSerCysArgLeuLeuGlyT1eGly 120  
Db 301 ATTCCCTTACCAAGGCTCTTGAAGGCTGCGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
QY 121 G1YH1sLeuSerProGlnGlyValaLeuThrLeuProGlnGlnH1sProCysLeuVala 140  
Db 361 GGGCACTTTTCCCAAGGAGGCACTCACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
QY 141 ProArgLeuArgCysGlnSerCysAlaLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 421 CCAAGGCTCGAGATCCAGTCTGCAAGCTGGCAAGGCGCCCGGGCTCAGGGGAGGAGGAGG 480  
QY 161 ArgSerProGlnGlnSerLeuGlyValaArgArgLeuValaAsnThrAlaProGln 180  
Db 481 AGGTCAACCAAGAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200  
Db 541 TTCAGGCCCCCACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600

QY 201 AlaSerLeuArgAlaLeuAspProAspGlnGlyValaGlyArgLeuGlnTyrThrMet 220  
Db 601 GCATCCCTGAGGGCCATCGACCCGGAGCGAGGGTGAAGGAGCTGACTGAGTACACAGT 660  
QY 221 AspAlaLeuPheAspSerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 240  
Db 661 GATGCCCTTGTGATAGCGCTCCCAACAGTCTTCTCCCTGAGCCAGTCACTGGTGA 720  
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlnThrLysSerThrHisValPheArgValThr 260  
Db 721 GTACCAACAGCGAGAGAGCTGATCTGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrLysLeuValThr 280  
Db 781 GGCACAGACCAAGGAGCTGAGCCCGAGCAAGAGTGGCTGGCTGACCTCACTCTGTTGTTACT 840  
QY 281 AspThrAspAspHisAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
Db 841 GACACCAATGACCAATGACCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 901 AACCTGAGGTTGGCTATGAGGCTCACTGTCAGGGCCAGCGATGATGCCCTCC 960  
QY 321 AsnAlaAsnI1eLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340  
Db 961 AATGCCAATATTCATGACCGCTGCTGAGGGGAGTCTGGGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
QY 341 GlnI1eAspProArgSerGlyValI1eArgThrArgGlyProValAspArgGlnGlnVal 360  
Db 1021 GAGATGACCTGCTGCTGGGAGTATCCAGACCGTGGCTGATGAGTGGAGAGAGGAGT 1080  
QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
Db 1081 GATCTTACACATGACGCTAGAGGAGTGAACAGAGTCAAGAGTCCGGAGTCTCGAGAT 1140  
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAspAspAspAlaProGlnPheSer 400  
Db 1141 ACCACAGCCGCTGTTTCTTCTTCTGAGAGATGACAAATGATATGCCCTCCAGATTAT 1200  
QY 401 GlnLysArgTyrValGlnValArgGlnAspValThrProGlyValaProValLeuArg 420  
Db 1201 GAGAGCGCTATGTGTCCAGGTGAGGAGAGTATGATCCAGGGGCCCACTGATCTCCA 1260  
QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValaHisTyrSerI1eMetSer 440  
Db 1261 GTCAAGCTCGGATCGAGACAGAGGAGCAATGCCGTGTGATCATATGATGATGATGATGATGAT 1320  
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAspValaLys 460  
Db 1321 GGCATGCTCGGGGACAGTTTATCTGATGTCGACAGCTGAGCTTGTGATGTGTGATGATGATGAT 1380  
QY 461 ProLeuAspTyrGlnThrThrLysGlnTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
Db 1381 CCTTTGACTATGAG 1440  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspI1eAsnAsp 500  
Db 1441 CGTCCCACTCTCTATATGCTCTGGCTTGGAGACGATCAAGTCTGTGATATCAAGAGAC 1500  
QY 501 AsnAlaProI1ePheValSerThrProPheGlnAlaThrValLeuGlnSerValaProLeu 520  
Db 1501 AATGCCCATATCTTCTGAG 1560  
QY 521 GlyTyrLeuValLeuHisValaGlnAlaI1eAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1561 GGCTTACTGTTCTCATATCAAGCTATCAAGCTATCAAGCTATCAAGCTATCAAGCTATCAAGCTAT 1620  
QY 541 GlnTyrArgLeuAlaGlyValaGlyHisAspPheProPheThrI1eAsnAsnGlyTyrGly 560  
Db 1621 GATATACGCTGCTGCTGGGAGCAATGATCTCCCTTCAATCAATGAGCAAGAGC 1680  
QY 561 TrpI1eSerValaI1eAlaGlnLeuAspArgGlnGlnValaAspPheTyrSerPheGlyVal 580

Db	1681	TGGATCTCTGGCTGCTGAACCTGAGCCGAGAGAACTGATTTCTACAGCTTGGGGTA	1740	Db	2761	TTTGACGAGATGACTTGAATGTTGTGGAGAGAAACAGCCCATTTGGGCTACCCGTG	2820
Qy	581	GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThrVal	600	Qy	941	AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetCysGlnIle	960
Db	1741	GAAAGCTCGAGACCATGGACCTCCAGACCTCACTGCTCGGGCAGTGTCAAGCTGACTGC	1800	Db	2821	GCCCGGGTCACAGACCATGACCCCGTAGAGGACCAATGCCAGATTATGACCAAGATT	2880
Qy	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluValThrValArgLeuAsn	620	Qy	961	ValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla	980
Db	1801	CTGGATGTCAACGACAAACATCCACCTTTACCACCAAGATACACAGTGGCGGCTCAAT	1860	Db	2881	GTTGAGGGCAACATCCCTGAGAGTCTTCCAGCTGGACATCTTCCGGGAGAGCTGACAGCC	2940
Qy	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640	Qy	981	LeuValAspLeuAspArgGluAspArgProGluValThrValIleGluAlaThrSer	1000
Db	1861	GAGATGCAAGCTGTGGGACACAGCGTGTGACGCTGACGCTGTGACGTAATGCTCAT	1920	Db	2941	CTGGATGACTTAGACTACAGAGACCGGCTTAGATAGCTCTGATCATCAGGCCACGTCA	3000
Qy	641	SerValIleThrThrCysGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660	Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	1921	ACTGTCATCACTGATCCAGATACACAGTGGCAATACCGAACCCTTCTCCATCAACAGC	1980	Db	3001	GCTCCTCTGTGTGAGCGGGGCTTACAGTCAAGTCCGCTCTTGAACGGAAATGACACCCA	3060
Qy	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluValArgGln	680	Qy	1021	ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040
Db	1981	CAAAAGTGTGTGGCTGTGATCCCTTGCCCTGGCCACTGACTACAACTTGAGCGGCGAG	2040	Db	3061	CCAGTGTGGGCAACTTTGAGATCCTTTCAACAACTATGTCAACAAATCGCTCAAGCAGC	3120
Qy	681	TyrValIleLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700	Qy	1041	PheProGlyGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	2041	TATGATGTGGCTGTATCCGCTCCAGATGGCACTCGGCAAGACACGCGACAGATTGGGTG	2100	Db	3121	TTCCCTGGGGGTGCATTTGGCCGAGTACCTGCCATGACCTCGATATGTCAGATATGCTG	3180
Qy	701	AsnValIThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720	Qy	1061	ThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080
Db	2101	AATGTCAACGAGACGCAACCCATGCTCTGTCTTCAAGCTCCCACTATACATGATGAT	2160	Db	3181	ACTTACAGCTTTGAGCGGGGAAATGAATCAAGCTGCTGCTCTCAAGCTCCAGGGGT	3240
Qy	721	ValAsnGluAspArgProAlaGlyThrThrValIleLeuIleSerAlaThrAspGluAsp	740	Qy	1081	GluLeuIysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	2161	GTTAATGAGGACCGGCGCGGACGACAGGTGTGTGATCAGCGCCACGATAGAGAGAC	2220	Db	3241	GAGCTGAAGCTAACCGCGGCACTGACAAACCGGCTCTGAGAGGCATATGAGCGGTG	3300
Qy	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760	Qy	1101	LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Db	2221	ACAGGTGAGAAATGCCGACATCACTTCAATGAGAGACAGATCCCAAGTTCCGCAATC	2280	Db	3301	CTGGGTGACAGCGGTACACAGCGGTACCGCCAGATGCGGTGCTGTACCATCATCTC	3360
Qy	761	AspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer	780	Qy	1121	ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg	1140
Db	2281	GATGACAGACACGGGGGCTGTCAACCCAGGCTGAGCTGACTAGAGACCAAGTGTCT	2340	Db	3361	ACCGATGAGATGTCTCACCCACAGATACGCTGGCGCTGAGAGCATTTCAACCGAGCGC	3420
Qy	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIysSerAspThrThrTyr	800	Qy	1141	PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Db	2341	TACACCCCTGGCCATTACTGCTCGGACAAATGGCATTTCCACAGAACTCCACATCTAC	2400	Db	3421	TTCCCTGTACACCATGTGACGCTCTTCAATCCAGCGGTGGCGCGCACCTGGCCACGCA	3480
Qy	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820	Qy	1161	ProAspHisValValAlaPheAsnValGluAspAspThrAspAlaProGluGlyHisIle	1180
Db	2401	CTGGAGATCTCTGTGAACGACGTGAATGACAAATGCCCTCACTCTCTGAGAGATCTTAC	2460	Db	3481	CCGAGCACAGTGTGTCTTCAAGCTACAGCGGAGACACCGACGCCCTCGGGGCGCACATC	3540
Qy	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840	Qy	1181	LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyGlyProProPheLeu	1200
Db	2461	CAGGCGACAGTCTATACAGATGTGGCAACCTTCACTAGCGTCTCTGACAGATCTCAACCACT	2520	Db	3541	CTCAACGTGAGCCTGTGGTGGGCGACGGCGCAAGGCCCGGGGCGGGCCCTTCTCTG	3600
Qy	841	AspArgAspSerGlyLeuAsnGlnArgValPheTyrThrPheGlnGlyGlyAspAspGly	860	Qy	1201	ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer	1220
Db	2521	GATCTGTGATTTCTGGACTTATATGGCAGGGTCTTCTTCACTTCCAGAGGCGCAATGTGA	2580	Db	3601	CCCTCTGAGGACCTTGACGAGAGCGCTTATACCTCAACCGCAACGCTGTGACCGGCATCTCG	3660
Qy	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880	Qy	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr	1240
Db	2581	GACGGTGAATCTTATGTGTATGATCAAGTCAAGGACATGTGCAACGCTACAGGAGGCTGAGT	2640	Db	3661	GCAACGCGGTGTGCTCTTGCAGACGACAACTGTGCTGCGGAGAGCCCTGCGAAGATCAC	3720
Qy	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro	900	Qy	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260
Db	2641	CGAGAGAACTGGCCACAGTATGTCTTGGGGCATATGACGTGGACAAAGGAGATGCCCA	2700	Db	3721	ATGGCGTGGGTGTGGTGGCTGTGAGTCTTCCGAGCTTCCGCGCTTCATCGCTCTCTCC	3780
Qy	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal	920	Qy	1261	ValLeuPheArgProIleHisAspProValGlyGlyLeuArgCysArgCysProProGlyPhe	1280
Db	2701	GCCCGCACCATATGAGATGACAGTCACTGTGTGATGTGAATGACAAATCCCCCTGTC	2760	Db	3781	GTCCTCTTCCGGCCCATACACCCCGTGGAGGGGTGGCTGCGCTGCGCCCGCGCTTC	3840
Qy	921	PheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal	940	Qy	1281	ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis	1300
				Db	3841	ACGGGTGACTACTGCGAGACCGAGGTGACCTTGTGTACTGCGCGGCTGTGGCCCCAC	3900

OY	1301	GLYHRCYSAAGSERARGLUGLYGLYTHYRCSLEUCYBAHASPGLYTYRTHGLY	1320
Db	3901	GGGCGCTGCCAGCCGGCAAGGGCGGCTACACCTGCCCTCTCTCGTGAAGGCTACACGGGT	3960
OY	1321	GLUHISCYSGIUVAlSERIALARGSERGLYHRCYSTRProGLYValCYGLYSANGLY	1340
Db	3961	GAGCACTGTAGGTAGTACGTCTCGCTACAGCCCTTGACACCCGGGGTCTTGCAGAAATGG	4020
OY	1341	GLYTHRCYSVALASPHEULEUVALGLYGLYPHELYSCYBAAPCYAPROSERGLYASPHE	1360
Db	4021	GGCACTGTGTCAACCTCGTGGGGCGGTTCAAGTGCAGATTGGCCACTGTGAGACTTC	4080
OY	1361	GLIUYSPROYTCYSGIUVAlThyThyRARGSERPHEPROALAHISERPHGLIETHRHE	1380
Db	4061	GAGAAAGCCCTACTGTCCAGGTATCAACGGCAAGCTTCCCGGCCACTCTTATCACTCTT	4140
OY	1381	ARGGLYLEUARGIUNARGPHEHISPHETHIRLEUALAUSERPHEALATHIRYGLIARG	1400
Db	4141	CGCGGCGTGGCGCAGCGTTTCCACTTCAACCTGGGCGCTCTGGTTTGCCACAAGAGAGCG	4200
OY	1401	ASPGLYLEULEULEUTYRASNGLYARGPHEANGLIULYSHISAPPHEVALALALEUGLU	1420
Db	4201	GACGGGTGTGCTGTGTATCAATGGCGTTTCAATGAGAGCAATGACTTGTGTGGCCTCGAG	4260
OY	1421	VALIRIEGINGIUNVALAGIULEUTHRPHESERISAGLYGUSERTHThyThyVALSER	1440
Db	4261	GTCATCCAGAGACAGCTGTACGTCACTTCTGTGCAAGGAGGTACCAACCAACGGGTGCC	4320
OY	1441	PROPHIVALProGLYGLYVALSERASPGLYGINTRPHISITHRVALGIMLEULYETTYR	1460
Db	4321	CCATTTCGTGCCCGAGAGAGTCAGTGATGGCCAGTGGCATTAAGGTACCTGAATACAC	4380
OY	1461	ASNTLYSPROLEULEUGLYGLINTHRGLYEURPROGINGLYPROSERGLIULYVALAL	1480
Db	4381	AATAAGCCACTGTGGGTGTCAAGAGGGCTCCCAACAGGCGCATCAGACAGAAAGGTGGCT	4440
OY	1481	VALVALTHIRVALASPGLYCYASAPTHRGLYVALALALEUATCPHGLYSERVALLEUGLY	1500
Db	4441	GTGTGTGACCGTGTGATGTGCTGTACACAGAGATGGCTTGGCTTGGATCTGTCTTGGGC	4500
OY	1501	ASNTYSPERCYVALAALAGIINGLYTHRGINGIYSEITLYLYSERLEUASPHEUTHR	1520
Db	4501	AACCTACTCTGTGGCTGCCACAGGACCCAGAGGTGGCAAGAAGTCTGTGATCTGACG	4560
OY	1521	GLYPROLEULEULEUGLYGLYVALPROASPHEUPROGLYUSERPHEPROVALARGMETARG	1540
Db	4561	GGGCCCCCTGTACTAAGGGCGGGTGTCACTGCCCGAGAGCTTCCAGTCCGAATGGCG	4620
OY	1541	GLINPHEVALGLYCYMETARGASNTLEUGINVALASPSEARGHISILASPMEVALASP	1560
Db	4621	CAGTTCGTGGGCTGTATGTGGAACTGTGACAGGTGTGACACACCGGCACTATGACATGGCTAC	4680
OY	1561	PHIELLEALASPANGLYTHYVALPROGLYCYSPROALALYLYASNVAlCYASPSER	1580
Db	4681	TTCAATTGCCAACAAATGAGCACCGGTGCTGTGCCCTGTCCCAAGAAAGAACTGTGTGTACAGC	4740
OY	1581	ASNTTHRCYSHISANGLYGLYTHRCYSHISANGLINTPRAPRALAPHESERCYGLIUCYS	1600
Db	4741	AACACTTGTCCCAAGGGGGGCACTTGTGTGAACCAAGTGGAGCGCTTCACTGTGCAAGTGC	4800
OY	1601	PROLEUGLYPHAGLYLYYSPERCYVALAGIINGIUMETALASPProGLIHISAPHELEU	1620
Db	4801	CCCTCGGGCTTTGGGGGCAAGAGCTGCCGCCCAAGAAATGGCCAAATCCACACACTTCTGTG	4860
OY	1621	GLYSPERLEUVALALATPRHISGLYLEUSERLEUPROIIUSERGINPROTPTRYLEU	1640
Db	4861	GGCAGACACCTGTGTGGCTGTGATGGCTCTGTGCCCAATCTCCCAACCTGTGATACCTC	4920
OY	1641	SERLEUMETPHEARGTHYRARGIINALASPGLYVALLEULEUGIINALIETHYRARGLY	1660
Db	4921	AGCTTCATGTTCCGACCGCGCCAGGCGCAAGGTGTCTGTGTGACAGGCATCAACAGGGGGG	4980

Qy	166	ArSeSerThrIleThrLeuGlnLeuAspArgGluGlyIleValMetLeuSerValGluGlyThr	1689
Dp	4881	CGCAGACCACTCAACCTTCAAGCTACAGAGGGCCACTGATGCTGAGCCGAGGGCACA	5040
Qy	1681	GlyLeuGlnIleAspSerLeuArgLeuGluProGlyValArgIleAsnAspGlyAspTrpHis	1700
Dp	5041	GGGCTTCAGGCTCCTCTCTCCGCTTGAGCCAGGCGGGCCCAATGACGGGATGTGGCAC	5100
Qy	1701	HisAlaGlnIleuAlaLeuGlyAlaSerGlyGlyProGlyHisValAlaIleLeuSerPheAsp	1720
Dp	5101	CATGACAGCTGGGACATGGGAGCCAGGGGGGGCGCTGGCCATGCATCTGCTCTTGAT	5160
Qy	1721	TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer	1740
Dp	5161	TATGGGACGACGAGGACGAGGGGCACCTGGGCCCCCGGCTGCATGCTGCACCTTGAGC	5220
Qy	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaIleArgGlyPheAspGly	1760
Dp	5221	AACCTTAACTATGGGCGGAAATACCTGGGCCAGCGCGCGTGTGGCCCTGTGCTTCGGGGC	5280
Qy	1761	CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer	1780
Dp	5281	TGTTTGCAGGGGTGTGGGGGTGAGCGGATACCCCAAGGGGGGTTAACAGCTTGGATCCAGC	5340
Qy	1781	HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
Dp	5341	CATGGGAGAGCACTCAACCTGGAGCAAGGCTGTAGCTCGCTCAACCTTGTGACTCAAC	5400
Qy	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp	1820
Dp	5401	CCGGTCTCTCTAACTAACTATGACAGAACGACTGGGAGCACTATCTCTCAAGCTGGAT	5460
Qy	1821	ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln	1840
Dp	5461	CAAGTTACTATGGTGAACACTGTAACATGTTGTGTAACCTGAACCCGTGTGAGCAACAG	5520
Qy	1841	SerValCysThrArgIlePheProSerIleProHisGlyTyrThrCysGluCysProProAsn	1860
Dp	5521	TCTGTGTGTACCCGACAGCCCAAGCGCCCCCAATGGCTATACCTGGCAGTGTCCCCCAAT	5580
Qy	1861	TyrLeuGlyProTyrCysGluThrArgIleAspGluAspProCysProArgGlyTyrTrpGly	1880
Dp	5581	TACCTTGGGCAATCTGTAGACCAAGATTGACCAAGCTTGTCCCTGGCTGTGGGGA	5640
Qy	1881	HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn	1900
Dp	5641	CATCCCACTATGGGCCCATGCAACTGATGATGACGAAGGCTTTGACCAAGACTGCAC	5700
Qy	1901	LeuThrSerGlyGlyCysHisCysGlyGluAsnHisGlyArgProProGlySerProThr	1920
Dp	5701	AAGCAAGAGCGGACAGTGCACCTGCAGAGAAACAACAACCAACGAGGCCCAAGGCCCAAC	5760
Qy	1921	CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp	1940
Dp	5761	TGCTCTTGTGTAGCTGTAACCCCAAGGCTCTCTTGTCCAGAGTCTGTGACCTTGAGAT	5820
Qy	1941	GlyGlnCysProCysIlePheProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro	1960
Dp	5821	GGCCAGAGTCAATGACAGAGCCAGAGTGCATGGGGCGTCAAGTGAACCGCTGTGACAACTT	5880
Qy	1961	PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle	1980
Dp	5881	TTTGTCTAGGTCACCAACCAATGGCTGTGAAGTAAATTATACAGCTGCCACAGAGCAATT	5940
Qy	1981	GluAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaIleProCysPro	2000
Dp	5941	GAGGCTGGATCTGTGTGGCCCCCGTCAACCCGCTGGGGCTGTCTGTCTCTCTCTGCC	6000
Qy	2001	LeuGlySerPheGlyThrAlaValArgHisCysAspGluHisValArgGlyTyrPheProPro	2020
Dp	6001	AAGAGCTCTTTTGGAGCTGCTGTGGCCCACTGTGATGACAGCAGGGGGGTGGCTCCCCCA	6060
Qy	2021	AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLeuGlyPheAlaGluArgLeu	2040

```

Db      6061 AACCTTTCAACTGCAAGTCATCACTTCTCAGAACTAAGGGCTTCCTGAGCGGCTA 6120
Qy      2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060
Db      6121 CAGCGGAATGAGTCAGGCTCAGACTCAGAGGCGCTCCAGAGAGTACCCCTGCTCGGCC 6180
Qy      2061 AamAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIleValAlaTyrGlnLeu 2080
Db      6181 AAGCGCACCGACACACACACTGCTACTTCGCGAGGAGCGTCAMAGGTGCGCTTACCGAGCTG 6240
Qy      2081 AlaThrArgLeuLeuAlaHisGlnSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db      6241 GCCACGCGGCTGCTGCGCCACAGAGACCCAGCGGGGCTTGGGCTGTCTGCACACAG 6300
Qy      2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db      6301 GACGTGCACTTCAGTCAAGATCTGCTGGGGTGGGAGGCGCTCTCTGAGCACAGCCAC 6360
Qy      2121 LysArgHisThrGlnLeuIleGlnGlnThrGlnGlyGlyThrAlaThrLeuGlnHis 2140
Db      6361 AAGCGGCACTGGAGGTGATCCAGACAGAGGGTGGCACCGCTGCTGCTCCAGCAC 6420
Qy      2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160
Db      6421 TAGAGGCTTACGCAAGTCCTGCGCCAGAACATGCGGCACTTAAAGCCCTTC 6480
Qy      2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspIleGlyAsnPheAla 2180
Db      6481 ACCATGTCAGCCCAACATTGTCATCTCGTAGTCCGCTTGGAACAAGGAATTGCT 6540
Qy      2181 GlyAlaIleLeuProArgTyrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200
Db      6541 GGGGCGCAAGCTGCGCGCTACAGAGCGCTGCGGGAGACAGCCCGGAGCTTGAACA 6600
Qy      2201 ThrValIleLeuProGlnSerValPheArgGlnThrProProValIleArgProAlaGly 2220
Db      6601 ACGATCATTTCTGCTGAGTCTGTCTTCAGAGAGAGCGCCCGGTGTCAGGCGCGAGGC 6660
Qy      2221 ProGlyGlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgArgHisPProGlnLeu 2240
Db      6661 CCCGAGAGGCGCCAGAGGCAAGAGAGCTGGCAGCGGCAAGAGAGGCAAGCGGAGGCTG 6720
Qy      2241 SerGlnGlyGlnAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260
Db      6721 AGCCAGGGTGAGGCTGTGGCCAGCGTCATCTACCGCACCTGCGCGCTACTGCTC 6780
Qy      2261 HisAsnTyrAspProAspIleArgSerLeuArgValProIleValIleAsnThr 2280
Db      6781 CATTACTTATGACCTTGACAGGCGACGCTTGAAGATCCCAAGCGCCGATCATCAACA 6840
Qy      2281 ProValIleSerIleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspIle 2300
Db      6841 CCCGAGGAGCATCAGGCTCATGATGAGAGGCTTCTGCGCGCGGCGCTGAGCAAA 6900
Qy      2301 ProValThrValGlnPheArgLeuLeuGlnThrGlnGlnArgThrIleProIleCysVal 2320
Db      6901 CCCGTCACGGTGCAGTTCGCGCTGAGACAGAGGAGCGGACCAAGGCCCATCTGTGTC 6960
Qy      2321 PheThrAsnHisSerIleLeuValSerGlyThrGlnGlyTyrPheSerAlaArgGlyCysGln 2340
Db      6961 TTCTGGAACATTCAATCTGTGTCAGTGCACAGGTGCTGCTGCGCCAGAGGCTGTAA 7020
Qy      2341 ValValPheArgAsnGlnSerHisValSerCysGlnCysAsnHisAsnMetThrSerPheAla 2360
Db      7021 GTCGCTTCCGCAATGAGAGCACCGTCAGCTGCAGTGCACCATGACGAGCTTCGCT 7080
Qy      2361 ValLeuMetAspValSerArgArgGlnAsnGlyGlnIleLeuProLeuIleThrLeuThr 2380
Db      7081 GTGCTCATAGGAGGTTTCTCGCGGAGAAATGGGAGATCTCTGCCATGAAGCACTGACA 7140
Qy      2381 TyrValAlaLeuGlnValThrThrLeuAlaIleLeuLeuThrPhePheLeuThrLeu 2400

```

---

```

Db      7141 TACGTGCTCAAGGATGTCACCTTGCGCTTCCTGCTCACCTTCTTCTCACTCTC 7200
Qy      2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleGlnGly 2420
Db      7201 TTGGGATATCTGCGCTCCACCAACAGCGCATCCAGATCCATGACAGCTGCTCGGGC 7260
Qy      2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db      7261 CTGGCTCAGCTGTCTTCTCTCGGAAATCAACAGGCTGACCTCCCTTCTTCGTCACA 7320
Qy      2441 ValIleIleAlaIleLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuGlnGlu 2460
Db      7321 GTCATTTGCATCTCTGTGACTCTCTGTACCTGTGCACTTTTCTGAGCTTGTGAGAG 7380
Qy      2461 AlaLeuHisIleLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480
Db      7381 GCGTTGCACTGTATACGGGCACTCATAGGTGGCGATGTCAACACCGGCCCATGCGC 7440
Qy      2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db      7441 TTCTACTACATGCTGGGCTGGGGCGTGCCTGCTTCATCAACAGGGCTAGCGTGGGCTG 7500
Qy      2501 AspProGlnGlyTyrGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuIle 2520
Db      7501 GACCCGAGGGCTACGGGAACTGACTTCTGTGCTCTCTCATCTATGACAACGCTCATC 7560
Qy      2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db      7561 TGGAGTTTGTGCGGCGCGGTGGCTTTCGCTTCGATGAGATGTTCTCTGTACATCTG 7620
Qy      2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnValGlyGlyProValSer 2560
Db      7621 GCGGCGCGGCGCTCTGTGCTGCCAGCGGAGGCTTGGAGAAAGTCTGTCTCG 7680
Qy      2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrThrLeuLeuAlaLeu 2580
Db      7681 GCGCTGACGCTCTCTTCTGCGCTCTCTGTGCTGAGCGGACAGCTGCTGCTGCACTG 7740
Qy      2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600
Db      7741 CTCTCTGTCAACAGAGAACCTCTCTTCACTTCACTTCTTGTGCACTGCAATTGCAATC 7800
Qy      2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerIleGlnValArgIleAlaLeu 2620
Db      7801 CAGGCGCCCTTCATCTTCTCTCTATGTGCTGTAGCAAGAGGTCCGGAAGAAGCACTC 7860
Qy      2621 LysLeuAlaCysSerArgIleProSerProAspProAlaLeuThrThrIleSerThrLeu 2640
Db      7861 AAGCTTGCTCAGCGCGGACCGGACCGCTGACCTGCTGTGACCAACAGTCCACCTG 7920
Qy      2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660
Db      7921 ACCTGCTCTTCAACTGCGCCAGCGCCCTTACCGCAATGGCGGCTGTACACACCTTACCG 7980
Qy      2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyIleSerGlnProSerTyrIle 2680
Db      7981 GACTCGCGCGCTCTGTGCAACAGCACAGTGTGCGGCAAGAGTACAGCCAGCTACATC 8040
Qy      2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlnGlnIleProProGlyLeuGly 2700
Db      8041 CCCTTCTGTGAGAGAGAGATCCGACCTGAACCTTGCGCAAGGCGCCCTGCGCTTGGGG 8100
Qy      2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720
Db      8101 GATCAGAGGAGCTGTTCTGTGAAGGTCAACAACAGAGCATGATCTCGACACCGGACTCC 8160
Qy      2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740
Db      8161 GACAGTGACCTGTCTTGAAGAGACACAGAGTGTCTCTATGTGCTTACCACTCACTACA 8220
Qy      2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760
Db      8221 GACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280

```

QY 2761 TTPASPSerLeuEnglyProglYAlaGluArgLeuProLeuHisSerThrProLyAsp 2780  
 Db 8281 TGGGATAGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8340  
 QY 2781 GAGGAGGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8400  
 Db 8341 GAGGAGGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8400  
 QY 2801 SerSerGlyAanglyAlaProgluGluArgLeuArgLysAnglyAlaSerArg 2820  
 Db 8401 AGTAGTGGCAAGGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8460  
 QY 2821 GluGlySerLeuEnglyProLeuProglYSerSerAlaGlnProHisLysGlyLysLeuLys 2840  
 Db 8461 GAGGGGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8520  
 QY 2841 LysLysSerLeuProThrLysSerGluLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
 Db 8521 AAGAGGTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8580  
 QY 2861 CysThrGlySerSerArgLysSerSerAlaSerGluLysSerArgLysGlyProProPro 2880  
 Db 8581 TGCAACAGGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8640  
 QY 2881 ArgProProProArgLysSerLeuGluGlnLysAnglyAlaMetProLysAlaMet 2900  
 Db 8641 CGCCCAAGGCTGTGGGGCCCTGGAGCAGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8700  
 QY 2901 SerLeuLysAlaGlyThrValAlaProGluLysSerSerGlyLysLeuPhePheAsn 2920  
 Db 8701 AGCATCAAGGAGGAGCGTGGATGAGAGACTGCCCTCGACAGTACTCCCAAGAT 8760  
 QY 2921 PheLeuHis 2923  
 Db 8761 TTCTCTGAT 8769

RESULT 2  
 AX549238  
 LOCUS AX549238 10531 bp DNA linear PAT 26-NOV-2002  
 DEFINITION Sequence 523 from Patent WO02061087.  
 ACCESSION AX549238  
 VERSION AX549238.1 GI:25813930  
 KEYWORDS  
 ORGANISM Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 CHORDATA; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 Burner, G.C., Roush, C.L. and Brown, J.P.  
 Antigenic peptides, such as for G protein-coupled receptors  
 (GPCRs), antibodies thereto, and systems for identifying such  
 antigenic peptides  
 Patent: WO 02061087-A 523 08-AUG-2002;  
 Lifespan Biosciences, Inc. (US)  
 FEATURES  
 SOURCE  
 1. 10531  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 2038 a 3353 c 2998 g 2142 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 10531  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-916-849a-3 (1-2923) x AX549238 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 Db 63 ATCGGAGCCCGGGCCACCGGGCGTCCCTCCCAACCGCGCGCGCTGTGTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuLysAnglyAlaGlyProCysArgSerLeu 40  
 Db 123 TTGCTGT 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSer 60  
 Db 183 GGGTCAAGGAGAGAGAGCTTTCGGGGGCTGTGGGGCCCAATGGGCTGTGTGTGTGTGT 242  
 QY 61 SerAlaSerAlaLeuThrLeuThrThrSerArgCysArgAlaGlyThrGluLeuThr 80  
 Db 243 TCAGGCTCAACCTGT 302  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuAlaValThrCysProGluSerGlyAlaHis 100  
 Db 303 GGCACCTGT 362  
 QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyLysGly 120  
 Db 363 ATTCCTCAACCACTGT 422  
 QY 121 GlyHisLeuSerProGluGlyLysLeuThrLeuProGluGluHisProCysLeuValAla 140  
 Db 423 GGCACCTGT 482  
 QY 141 ProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
 Db 483 CACGGCTCAAGAGCAGT 542  
 QY 161 ArgSerProGluGluSerLeuGlyArgArgLysArgAsnValAlaThrAlaProGln 180  
 Db 543 AGGTCAACAGAGAGTCCCTGT 602  
 QY 181 PheGlnProProSerGlyThrAlaThrValProGluLysGlnProAlaGlyThrProVal 200  
 Db 603 TTCCACGCTGT 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyValArgLeuGluThrMet 220  
 Db 663 GCATCTGT 722  
 QY 221 AspAlaLeuPheAspSerArgSerArgSerAlaPheSerLeuAspProValThrGlyAla 240  
 Db 723 GATGCTCTGT 782  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
 Db 783 GTAACCAACGCGCAGAGAGT 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrLysLeuValThr 280  
 Db 843 GCGCAGAGCAGGAGATGCCCGCAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 902  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGluGluThrLysGluSerLeuArgLys 300  
 Db 903 GACACCAATGACCATGACCTGT 962  
 QY 301 AsnLeuGluValGlyThrGluValLeuThrValArgAlaThrArgAspLysAlaProPro 320  
 Db 963 AACCTGAGGT 1022  
 QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGluGlySerGlyLysProSerGluValPhe 340  
 Db 1023 AATGCCAATTTTGT 1082  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgLysGluVal 360  
 Db 1083 GAGATGACCTGT 1142  
 QY 361 GluSerTyGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380



```

Db      1143 GAATCTTACAGCTTACGCTGAGGCAAGTACCAAGGCTGGGACCCGGCTCTCGAAGT 1202
Qy      381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPheSer 400
Db      1203 ACCACAGCCGCTGTTTCTCTTCTGCGAGGATGACAAATGATATGCCCCCGAGTTTAAGT 1262
Qy      401 GluLysArgTyrValValGlnValArgGluAspValThrProGluValAlaProValLeuArg 420
Db      1263 GAGAACCGCTATATGTGTCCAGGTGAGGAGATGTGATCTCCAGGGGCCCCAGATCTCCAA 1322
Qy      421 ValThrAlaSerAspArgAspLysArgLysSerAsnAlaValValHisTyrSerIleMetSer 440
Db      1323 GTCAACGCTTCGATGACGACAGGGAGCAATGCGGTGTGATCATATGACATCATAGAGT 1382
Qy      441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460
Db      1383 GCGAATGCTCGGGGACAGTTTATCTGGATGCCCAAGCTGGAGCTCTGGATGTGGAGAC 1442
Qy      461 ProLeuAspTyrGluThrThrTyrGlyGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db      1443 CCTCTTGACTATATGACGACCAAGAGATCACCTCTAGGGGTGGACACAGATGTGGGC 1502
Qy      481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db      1503 CGTCCCTCACTCTATATGCTCTGGCTGTGGTACATAGGCTCTGGATCATACAGAC 1562
Qy      501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      1563 AATGCCCCCATCTTCGTCAAGACCCCTTTCAGAGCTACTGTCTGTGAGAGCGGCCCTTA 1622
Qy      521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db      1623 GCGTACCTGGTCTCTCATGTCCAGGCTATCGACGCTATGCTGTGACATGTGCCCTCG 1682
Qy      541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1683 GAATACCGCTTGGCTGGGTGGACATGACTTCCCTTCAACATCAATGACACAGGC 1742
Qy      561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      1743 TGGATCTCTGTGGCTCTGACTGACGACCGGGAGGAGTTGATTCTTCAAGCTTTGGCGTA 1802
Qy      581 GluAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      1803 GAAGCTCGAGACCATGTGCACTCCAGCACTCACTGCTCGGCCAGTGTCAAGCTGTCTGC 1862
Qy      601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      1863 CTGGATGTCAAGCAACAATCCAACTTTTACCACAGAGTACACAGTGGCGGCTCAAT 1922
Qy      621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      1923 GAGGATGCGAGCTGTGGGACACAGCGTGGTGAAGGTGTCAAGCTGTGACCGTATGTCTAT 1982
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1983 AGGTGCATCACTTACAGATTCACACAGTGGCAATACTCGAAACCGCTTCTTCATTCACAGC 2042
Qy      661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnArgGln 680
Db      2043 CAAAGTGTGTGGTGGCTGTGATTCCTTGGCTGGCCCTGACCTGACCTTAACAATGAGCGGAG 2102
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyTyrArgGlnAspThrAlaGlnIleValVal 700
Db      2103 TATGTGTGTGGCTGTATCCGCTCCGATGGCACTCGGACAGACACCGCACAGATGTGGTG 2162
Qy      701 AsnValAlaThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrTyrValAsn 720
Db      2163 AATGTACACGAGCGCAACCACTGTCTGTCTTCAAGCTCCCACTATACAGGAAT 2222
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740

```

---

```

Db      2223 GTTATATAGACACCGGCGCGACGACCAACGCTGTGTCTGATACGCCCAAGATGAGAC 2282
Qy      741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2283 ACAGGTGAGAAATGCCGATATCACTTACTTCAAGAGACAGATATCCCGAGTTCCGCAATC 2342
Qy      761 AspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2343 GATGCAAGACACGGGGGCTGTCAACCCAGCGCTGAGTGTGATCAAGAAACCAAGTGTCT 2402
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyr 800
Db      2403 TACACCTTGGCCATTAATGCTCTGGGACAAATGGCATTTCCCAAGAGTCCGACACACTAC 2462
Qy      801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db      2463 CTGGAGATCTGTGTGAAGAGATGTGCACCTTCACTAGCGTCTGTGACAGATCTCAGCACT 2522
Qy      821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2523 CAGGCAAGTGTATATGAGATGTGCACCTTCACTAGCGTCTGTGACAGATCTCAGCACT 2582
Qy      841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db      2583 GATCGTATTTCTGACTTAATGTGCAAGGCTTCTTACACTTCAAGAGGCGACAGATGSA 2642
Qy      861 AspGlyAspPheIleValGlySerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db      2643 GACGGTACTTATATTGTTGATGTTCAGGCACTGTGCGCACTGTGCGGAGCTTACGGAGCTGAGT 2702
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900
Db      2703 CGAGGAACGTGGCCCACTATGTCTTGGCGGATATGTGAGTGTGACAAAGGGATGTGCCCA 2762
Qy      901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db      2763 GCCCGCACTATATGAAATGACATGCTACTGTGTGATGTGAATGTGAACAATCCCTGTTC 2822
Qy      921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2823 TTATGAGCAGATGATGTTGATGTGTTGTGGAAAGAAACAGCCCATTTGGGCTTACCGGTG 2882
Qy      941 AlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      2883 GCCCGGGTCAACAGCCATGACCCCGCATGAAAGGCAACAATGCCCAATATATGTACAGATT 2942
Qy      961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db      2943 GTGGAGGCAACATCTCTGAGGTCTTCCAGCTGACATCTTCTCCGGGAGCTTGACAGCC 3002
Qy      981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db      3003 CTGGTATGACTTATGACATGAGAGACCGGCTGATGCTGTGCTGATCATCAGGCCAGTCA 3062
Qy      1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020
Db      3063 GCTCTCTGTGGAGCGGGGCTTACATGTCAAGCTCCCTTGTGACCGGCAATGACAAACCA 3122
Qy      1021 ProValIleuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040
Db      3123 CCAATGCTGGGCACTTTGATGATCTTTTCAACAATATGTACCAATGCTCAACAGCAC 3182
Qy      1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3183 TTCCCTGGGGGTGCATTTGGCCAGATGACTGCGCCATGACCTGATATCTCAGATGTCTG 3242
Qy      1061 ThrTyrSerPheGlnArgGlyAsnGluLeuSerLeuValIleLeuAsnAlaSerThrGly 1080
Db      3243 ACTTACAGCTTTGAGCGGGGAAATGAATCTGAGCTGTGCTCAATGCTCCACAGGGGT 3302
Qy      1081 GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db      3303 GAGCTGAAGCTTAAAGCCGCGCACTGAGCAACAACCGGCTCTGTGAGGCAATATGAGCGTG 3362

```



QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
DB 3363 CTGGGTCTCAGACGGCGTACACAGCGTACCGCCAGCGCGCTGCGTGTGACCATCATC 3422  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
DB 3423 ACCGATGAGATGCTCACCCACACATCACTGCGCTGTGAGGACATGTCAACCGAGCGC 3482  
QY 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaThrLeuAlaThrPro 1160  
DB 3483 TTCTGTGACCACTGCTAGGCGCTTTCATTCAGCGCGGTGGCCGCCACGCTGGCCAGCCA 3542  
QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
DB 3543 CCGGACCACTGGTGTCTTCAACGTACAGCGGACACCGACCGCCCGGCGGCGCACATC 3602  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPheLeu 1200  
DB 3603 CTCAAGTGAAGCTGTGCGTGGGCGAGCGCCAGCGCCCGGCGCGCCCTTCCTG 3662  
QY 1201 ProSerGluAspLeuGlnArgLeuYrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
DB 3663 CCTTGTAGGACCTGACAGAGCGCGCTTATCTTCAACCGAGCGCTGTGACGCGCATCTG 3722  
QY 1221 AlAGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnYr 1240  
DB 3723 GCACAGCGCTGCTGCGCTTGCAGACACATCTGCTGCGGAGCGCTTGCAGAACTAC 3782  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1260  
DB 3783 ATGCGCTGCGTGTGCGTGGCGCTTGCATCTTCCGCGCGCTTCACTGCTCTCTCTC 3842  
QY 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
DB 3843 GTGCTCTTCCGCGCCATCCACCGGTGAGGGGTGGCGTGGCGCGCGCGCGCGCTTC 3902  
QY 1281 ThrGlyAspYrCysGluThrGluValAspLeuCysYrSerArgProCysGlyProHis 1300  
DB 3903 ACCGGTGACTACCTGCGAGCGAGGAGTGAGACTCTGCTACCTGCGCGCTGTGCGCCAC 3962  
QY 1301 GlyArgCysArgSerArgGluGlyYrThrThrCysLeuCysArgAspGlyYrThrGly 1320  
DB 3963 GGGCGCTGCGCGAGCGCGAGGGGCGCTACACCTGCTGTGTGTGTGTGTGTGTGTGTGT 4022  
QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly 1340  
DB 4023 GAGCACTGTAGGGTGTGCTGCTGCGCGCTTGCACCCCGGTGTCTGTGAGAGATGG 4082  
QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysArgCysProSerGlyAspPhe 1360  
DB 4083 GGCACTGTGTCAACCTGTGTGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4142  
QY 1361 GlnLeuProYrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
DB 4143 GAGAGCGCTTACTGCGAGGTGACACGCGAGCTTCCCGCCACTCTTCACTCACTT 4202  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrGlyGlyArg 1400  
DB 4203 CGGGCGCTGCGCGAGCGCTTCACTTCACTTCCCTGTGTGTGTGTGTGTGTGTGTGTGT 4262  
QY 1401 AspGlyLeuLeuLeuYrAsnGlyArgPheAsnGlyLeuHisAspPheValAlaLeuGlu 1420  
DB 4263 GACGGGT 4322  
QY 1421 ValIleGlnGlnValAlaGlnLeuThrPheSerAlaGlyGluSerThrThrThrValSer 1440  
DB 4323 GTGATTCAGAGAGAGTCAAGTCACTTCTCTGCGAGGAGTCAACACAGAGTGTCC 4382  
QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheThrValGlnLeuYrYrYr 1460  
DB 4383 CCAATTCTGCGCGAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4442

QY 1461 AsnYrProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLeuValAla 1480  
DB 4443 AATTAGCCACTGTGTGGTCAAGAGGCTCCACAGAGGCGCATCAGAGAGAGAGTGGCT 4502  
QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
DB 4503 GTGGTGAACCGGTGAGT 4562  
QY 1501 AsnYrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLeuYrSerLeuAspLeuThr 1520  
DB 4563 AACTACTCTGT 4622  
QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
DB 4623 GGGCGCGCTGTACTAGGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4682  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisAlaAspMetAlaAsp 1560  
DB 4683 CAGT 4742  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLeuYrAsnValCysAspSer 1580  
DB 4743 TTCAATTGCCAACAATGTGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4802  
QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
DB 4803 AACCTTGTGCACAATGTGGGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4862  
QY 1601 ProLeuGlyPheGlyGlyCysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
DB 4863 CCCCTGGGCTTTGGGGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4922  
QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProTrpYrLeu 1640  
DB 4923 GGACAGAGCTGT 4982  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
DB 4983 AGCCTCAATGTTCCGACCGCGCAGGCGCACGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5042  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
DB 5043 CGCAGACCACTACCTTACGCTTACGAGAGGCGCACGTGATCTGAGCGTGAAGGCGACA 5102  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrpHis 1700  
DB 5103 GGGCTTACAGGCT 5162  
QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaAlaLeuSerPheAsp 1720  
DB 5163 CATGACACAGCTGTCACTGGAGACAGCGGGGGCGTGTGTGTGTGTGTGTGTGTGTGTGT 5222  
QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleLeuSer 1740  
DB 5223 TATGGCAGACAGAGAGAGAGGCGAAGCTGTGGCGCGCGGTGTGTGTGTGTGTGTGTGTGT 5282  
QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
DB 5283 AACATTAAGTGGGCGGAATACCTGGGCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 5342  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAlaAsnSerLeuAspProSer 1780  
DB 5343 TGTTTGACAGGT 5402  
QY 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
DB 5403 CATGGGAGAGGATCACTGTGAGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5462  
QY 1801 ProCysProAlaAsnSerYrCysSerAsnAspTrpAspSerYrSerCysSerCysAsp 1820  
DB 5463 CCGTGTCTGTAAAGCTAATGTGACAGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5522  
QY 1821 ProGlyYrYrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlyGln 1840

```
Db 5523 CGAGTTACTATAGTGAACACTGACTAATGTGTGTAACCTGAACCCGTGTGAGCAAG 5582
Qy 1841 SerValCysThrArgLysProSerAlaProHisglYThrCysgluCybProProAsn 1860
Db 5583 TCTGTGTATCCCGCAAGCCCAAGTGCCTCCCAAGGCTATACCTGCGAGTGTCCCAAT 5642
Qy 1861 TyrLeuGlyProTyrCysgluThrArgIleAspGlnProCybProArgIlyTTPTrgLy 1880
Db 5643 TACCTTGCGGCATATCTGTAGACCAAGGATTGACCAAGCTTGTCCCGGTGCTGTGGGA 5702
Qy 1881 HisProThrCysglYProCybAsnCybAspValSerIlysglyPheAspProAspCybAsn 1900
Db 5703 CATCCACAGTGTGCCCATCATGCACTGTGATGTGACGAAGGCTTTGACCCAGACTGCAAC 5762
Qy 1901 LysThrSerIlygluCybHisCysIlysgluAsnHisThrArgProProglYSerProThr 1920
Db 5763 AAGACAAAGCGGCAAGTCCACTGCAAGAGAACCACTACGGGCCCCAGGCAAGCCCAACC 5822
Qy 1921 CysLeuLeuCybAspCybTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940
Db 5823 TGCCTCTGTGTGACTGTACCTCCCAAGGCTCTTGTCCAGAGTCTGTGACCTTGAGGANT 5882
Qy 1941 GlyGlnCybProCybLysProGlyValIleGlyArgGlnCybAspArgCybAspAsnPro 1960
Db 5883 GGCACAGTGTCAATGCAAGCAAGGTGTCTATGCGGCTGCAAGTGTACCGCTGTGCAACCT 5942
Qy 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProAspAlaIle 1980
Db 5943 TTTCGTAGAGTCAACCAATGCTGTGAGTGAATTAATACAGCTGACCCCAAGAGAAAT 6002
Qy 1981 GluAlaGlyIleTTPTrpProArgThrArgPheGlyLeuProAlaIleAlaProCybPro 2000
Db 6003 GAGGCTGGAGATCTGTGGCCCTGATCCCGCTTGAGGCTGCTGCTGCTGCTGCTGCTG 6062
Qy 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArgGlyTTPLeuProPro 2020
Db 6063 AAGGCTCTCTTGTGGACTGTGTGCGCACTGTGATGAGACAGGGGTGGCTCCCA 6122
Qy 2021 AsnLeuPheAsnCybThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu 2040
Db 6123 AACCTTTCACTGCAAGTCACTCACTTCTCAAGACTGAAGGCTTCCTGTAAGCGGCTA 6182
Qy 2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060
Db 6183 CAGCGGAATGAGTCAAGGCTTGAAGTCAAGGCGCTCCAGCAAGCTAACCTGCTGCTG 6242
Qy 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIlyValAlaTyrGlnLeu 2080
Db 6243 AAGCCACGAGACACAGCTGTGCTTCTGCGAGGAGCGTCAAGGTGCGCTAACGCTG 6302
Qy 2081 AlaThrArgLeuLeuAlaHisgluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db 6303 GCCACCGGCTGTGCGCCACAGAGACACCCAGCGGGCTTTGGGCTGTCTGCCACAG 6362
Qy 2101 AspValHisPheThrGlnLeuLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db 6363 GACGTGCACCTTCACTGAGATCTGCTGGGAGTGGAGCGGCTCTCTGAGACAGCAAC 6422
Qy 2121 LysArgHisThrGlnLeuLysGlnGlnThrGlnGlnGlyThrAlaTTPLeuGlnHis 2140
Db 6423 AAGCGGCACTGGAGGTGATTCAGAGACAGAGGGTGGACCGCTGTGCTCCAGCAC 6482
Qy 2141 TyrGluAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160
Db 6483 TATGAGCGCTAAGCGAGTGGCTGGCCCGCAAGACATGCGGCAACCTTAAGCCCTTC 6542
Qy 2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspIlysglyAsnPheAla 2180
Db 6543 ACCATGTGTACGCGCAATGTGCATCTCGTAGTGCCTGTGACAAAGGAACTTGTGCT 6602
Qy 2181 GlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAspLeuGlnThr 2200
Db 6603 GGGGCCAAGCTGCCCCGCTACGAGGCTTCGTTGGGAGACAGCCCCGCACTTGAGACA 6662
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220
Db 6663 AAGCATTTCTGTGATGTGTCTTCAAGAGAGACCCCCGTGTGACGGCCGAGGC 6722
Qy 2221 ProGlyGluAlaGlnGluProGluGlnLeuAlaArgArgGlnArgArgHisPProGluLeu 2240
Db 6723 CCGGAGAGGCCAGGAGCCAGAGAGAGCTGCAAGCGGCAAGGAGCGCACCCGAGCTG 6782
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrIleuAlaGlyLeuLeuPro 2260
Db 6783 AGCCAGGTTGAGGCTGTGCGCAGCGTCACTCATCTACCGCACCTGTGCGGGCTAGCTGCT 6842
Qy 2261 HisAsnTyrAspProAspIlyAspSerLeuArgValProLysArgProIleIleAsnThr 2280
Db 6843 CATTAATTAAGCCCTGACAGGCGAGCTTGAGAGTCCCAAGGCCCATCATCAACA 6902
Qy 2281 ProValValSerIleSerValHisAspAspGluGluLeuProArgAlaLeuAspLys 2300
Db 6903 CCGTGTGAGCATCAGGCTCCATGATGATGAGAGCTTCTGCCCCGGGCCCTGAGCAAA 6962
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGlnGluArgThrIlyProIleCysVal 2320
Db 6963 CCGGTCAAGGTGACAGTTCGCTGCGTGGAGACAGAGGCGGACCAAGCCCATCTGTGTC 7022
Qy 2321 PheTPAsnHisSerIleLeuValSerGlyThrGlnGlyTTPSerAlaArgIlyCysGlu 2340
Db 7023 TTCTGAACTTCAATTCATCTGATGAGTGCAGAGGTGTGTGTGCGGCAAGAGCTGTGA 7082
Qy 2341 ValValPheArgAsnGlnSerHisValSerCysGlnCysAsnHisAmcThrSerPheAla 2360
Db 7083 GTGCTTCCGCAAGAGAGGCCAGCTCAGCTGCCAGTGCACATGACAGACTTCGCT 7142
Qy 2361 ValLeuMetAspValSerArgArgGluLeuGlnIlyLeuProLeuLysThrIleuThr 2380
Db 7143 GTGCTCAAGAGTGTCTCGGGGAGAAATGGGAGATCTGCTGCATGAAGACATGACA 7202
Qy 2381 TyrValAlaLeuGlyValThrIleuAlaLeuLeuLeuThrPhePheLeuThrIleu 2400
Db 7203 TACGTGGCTTGAAGTGTCACTTGGCTGCTTGTGCTCACTTCTTCTTCTCACTCTC 7262
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisglYIleArgArgAsnLeuThrAlaAlaLeuGly 2420
Db 7263 TTGCGATTCCTGCGTCCAAACCAACAGGCAATCCAGCTAATCTGACAGCTCCCTGGGC 7322
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db 7323 CTGGCTCAGCTGTCTTCTCTCGGAAATCAACAGGCTGAACCTCTTGTGCTGACA 7382
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCybThrPheSerTPAlaLeuLeuGlu 2460
Db 7383 GTCAATTGCCATCTCTGTCACTTCTGTACCTGTCACTTCTTCTGGCTGTGCGAG 7442
Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480
Db 7443 GCGTTGACACTGTATCCGGGCACTCACTGAGGAGCGCATGTCAACACCGGCCCATGGGC 7502
Qy 2481 PheTyrTyrMetLeuGlyTTPGlyValIProAlaPheIleThrGlyLeuAlaValIlyLeu 2500
Db 7503 TTCTACTACATGTGAGGCTGGGGCTGTGCTGTTCATCAACAGGCTAGCCGTGGGCTG 7562
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTTPLeuSerIleTyrAspThrIleuIle 2520
Db 7563 GACCCCGAGGCTACGGAAACCTTGACTTCTGTGCTGTCTCATCTTAAGACAGCTCATTC 7622
Qy 2521 TTPSerPheAlaGlyIProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db 7623 TGAAGTTTGTGCGCCCGGTGGCTTTGCGCTGCAAGTGAAGTCTTCTGTGAATCTGCTG 7682
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLysIlysglyIProValSer 2560
Db 7683 GCGGCCCGGCTCTGTGTGCGCCAGCGGAGGCTTTGAGAAAGAGTCTGTCTGCTG 7742
```

Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
 Db 7743 GGCTGGAGGCTCTCTTGGCGTCTCTCTGCTGAGGCGCAGTGGCTGCTGCACTG 7802  
 Qy 2581 LeuSerValAnSerAspThrLeuLeuPheH1sTryLeuPheAlaThrCysAsnCysIle 2600  
 Db 7803 CTCTCTCTCAACACGACGACCTCTCTCTCACTCACTCTCTCTCTCTCTCTCTCTCT 7862  
 Qy 2601 GlnGlyProPheH1sPheLeuSerTrpValValLeuSerTrpGlyValArgValAlaLeu 2620  
 Db 7863 CAGGAGGCT 7922  
 Qy 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrTrpLysSerThrLeu 2640  
 Db 7923 AAGCTGCTGACAGCCGAGAGCCAGGCTCTGACCTGCTGACACCAAGTCCACCTG 7982  
 Qy 2641 ThrSerSerTrpAnCysProSerProTrpAlaAspGlyValGlyLeuTrpGlyProGly 2660  
 Db 7983 ACCCTGCTTCAACACTGCGCCAGCCCTTACAGATGGCGCTGTACCGAGCCCTACGGA 8042  
 Qy 2661 AspSerAlaGlySerLeuH1sSerThrSerArgSerGlyLysSerGlnProSerTrpIle 2680  
 Db 8043 GACTCGGCGGCTCTCTGACACGACAGTCTCTGCGCAAGAGTCAAGCCAGCTTACATC 8102  
 Qy 2681 ProPheLeuLeuArgLysLysSerAlaLeuAnProGlyGlnGlyProProGlyLeuGly 2700  
 Db 8103 CCTCTTCTGCTGAGGAGGAGTCCGCACTGAACTGCGCCAGGAGCCCTGCGCTGGGG 8162  
 Qy 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnH1sAspProAspThrAspSer 2720  
 Db 8163 GATCAGAGCGAGCTGCTCTGGAAGTCAACAGCAGCAGTGTCTCTGACACCGAGCTCC 8222  
 Qy 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTrpAlaSerThrH1sSerSer 2740  
 Db 8223 GACCTGACCTGCTCTTGAAGACGACCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 8282  
 Qy 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 Db 8283 GACAGTGAAG 8342  
 Qy 2761 TrpAspSerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2780  
 Db 8343 TGGGATATGCTGCTGAGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8402  
 Qy 2781 GlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 2800  
 Db 8403 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8462  
 Qy 2801 SerSerGlyAsnGlyAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2820  
 Db 8463 AGTAGTGGCAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8522  
 Qy 2821 GlnGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2840  
 Db 8523 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8582  
 Qy 2841 LysLeuPheCysLeuProThrIleSerGlnLysSerSerLeuLeuGlnGlnGlnGlnGln 2860  
 Db 8583 AAGAGAGTCTGCTGACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642  
 Qy 2861 CysThrGlySerSerArgLysSerSerAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGln 2880  
 Db 8643 TGCAAGAGGCT 8702  
 Qy 2881 ArgProProArgGlnInsLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2900  
 Db 8703 CGCCACCGCCCGGAGAGGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8762  
 Qy 2901 SerIleLysAlaGlyThrValAspGlnLysSerSerGlySerGlnPheLeuPhePheAsn 2920  
 Db 8763 AGCATCAAG 8822

Qy 2921 PheLeuH1s 2923  
 Db 8823 TTCCTGCAT 8831  
 RESULT 3  
 AF234887  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 10531)  
 Vincent, J.B., Skaug, J. and Scherer, S.W.  
 The human homologue of flamingo, *Scfz*, encodes a brain-expressed  
 large cadherin-like protein with epidermal growth factor-like  
 domains, and maps to chromosome 1p13.3-p21.1  
 DNA Res. 7 (3), 233-235 (2000)  
 MEDLINE  
 PUBMED  
 10907856  
 2 (bases 1 to 10531)  
 Vincent, J.B., Skaug, J. and Scherer, S.W.  
 Direct Submission  
 Submitted (15-FEB-2000) Department of Genetics, The Hospital for  
 Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada  
 location/Qualifiers  
 1..10531  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p21.1"  
 /clone\_lib="Stratagene fetal brain library, Cat. number  
 H13003a"  
 63..8834  
 /note="FMI; brain-specific, cadherin-like protein; similar  
 to the Homo sapiens product encoded by clone KIA0279 in  
 GenBank Accession Number D87469 and Rattus norvegicus  
 MGF2 encoded by GenBank Accession Number AB011528"  
 /codon\_start=1  
 /product="FLAMINGO 1"  
 /protein\_id="AA000080.1"  
 /db\_xref="GI:9828190"  
 /translation="MSSPATGVPLPPPLILILILILPPPLIGVQVPCRSRG  
 KSSSGACAMWGLCPSSASNLMTYRCKDAGTEULGHLVPHHDGRVVCPSREAHIP  
 LPPAPGCPWSCRLDGLIGHLSPQGLTPPEHPCLAKPRLCQSCKLQAPGLRAGE  
 RSPBSLIGRRKRNVTAPQFPSPYQATVPENQAGTVPASIRALDPDEGEARLEY  
 TMDLFDSSNQPFSLDPYTGAVTAELEDRKSTHVRVTADQHGMPRRSLATLT  
 ILVTNDNDPVEBOQYKSLRENLVEYELVTATGDGAPNNIILYRLLEGSGG  
 SESEVEIDPREGVTRGPDREEREYVLTASDQDRDGRSTAAVPLFSVD  
 NDNAIPSEKRYVQVREYVTPGAPLTATASDRDQGSNAVHTGISMSNAQPFID  
 AQTGALDVSPGLDYEETTKSLKVRQDGERPLSNVSGLVTVQVLDINDAFIVST  
 PFGATLVESPLDYLTKSLKVRQDGERPLSNVSGLVTVQVLDINDAFIVST  
 ELDEREVDYSGVEARHDGTPALTASASVSVLVLDVNDNPPFTQPEYVRLMBEA  
 VGSVTVVAVDRDAHSVITVQTSNTNRPSITSGSGGLVSLPLDYKLERVY  
 LATTASDGRDPTAOIIVAVTANTHRPVFQSHYVYNNEDRPAGTIVLLSATSD  
 TGENAIITTPMEDSIPQRIADDTGAVTQABLDYDQVSYTLATARPNGIPKSDT  
 TTEILVNDVNDAPQPLDSTQGSYEDVPPFTSVLQISATDRDGLNARVYTTQG  
 GDGDGDFTVESVSGIVRLRLDRENVQVLAAYADKMPATPMEVTVVLDV  
 NDNPVFBQDEPDPVVBENSPIGLAVARATADPDGTNAQIYQIVEGNIPEVQID  
 IFAGELTALVLDYEDRPEVAVTQATSAPLVSRAVHVALDRNDNPPVLAGFEILF  
 NNVTNRSSSPFGAIGRPADHPDSDSLTYSFEEGNEISLVLANSTGLSLAL  
 DNRPPLAIVSVLVDSVSVTAQCLATVITLIDELTHSITLRDMSBRLSLD  
 GLPTQAVATLTTPDPHYVTVNQRTDAPGHIILNVSLSVQPPPGGGRPTLPSD  
 LGRLTLNLSLTLAISAQVLPDPDNICTAREPENTMKCSVAFPSASAFILSSSV  
 FRILHPVGRIRCGCPGFTADYCEITVDICTSPGPHGRCSRBGGYTLCLDGDTG  
 EHEEVSARGRCIPGVCKNGGTCVNLVSGFKDCSDGPFKEKYCVTTKSPFASHT  
 TFRGLRORHFTLALSFAIKERDGLLVNGRFEKDKDPAVALVIOVOVLTFSAGST  
 TVYSPVPGVSDGWHVQLKYNNKPIILGGGILPGGSPBOKVAVTVVGGCDTGVALR  
 FGSVLNYSCAAQGTGSSKSLDLTGLLIGVPPDPSFPVRMQPFCGRANLQVD

SRHIMADPIANNGTVPCGPARKVCDSTNGCTGVNQMDAFSCCEPLGFGKSCA  
 QEMAPROFTLSSGLVAMHGLSLPISQWYLSLMPFRQADGVLLQATIGRSTLTLOL  
 REGHVMLSVEGTGLQASSLRLEPGRANDMDHQAOLAGAGSGPHALISLPGQORA  
 BENLGPRLHGLHLSNTVGGIIPGAGVAGRGSCIGRVADTEGVEGSLDPHSGES  
 INVEQCSLIPDCSDNPGKANGYCSNDMSYSGSCDGYGDNCTNVCDINCEHOSV  
 CTRKSGAPHGYCECPNYLIGPYCETRIDPCPGMMHGCPGKPCNOCVSKSPDDCN  
 KTSRGCKENHTRPPGSPCTCLCDCTPTGSLSRVCDPBDGCCPKPGVIGRCORCD  
 NPFAVYTNCEVNDSCPRALBAGIMPRTRFGLPAAPCKSGFGLVIRQDBERG  
 WLPMLNCTLSITFSELKGFARLQNMESGLSGSOQALALRNATHTAGYFGSDV  
 KYAYOLATRLHLSFSGRFGLSATODVFTENLRVSGALIDTANKHWEIIOQTEG  
 GTAMLOHYEAYASALQNMRTYLSPTITVPTNIVISVRLDKNHPAGAKLPRYAL  
 RSGOPDLETTVILPESVRETPVVRPGBOBPBEPALROKHNRHPLSGEZYASV  
 VIVRTLGLALPHANDPDRSRARVGRPIINTPVVISIVHDBELLPRALDPVTVOF  
 RLLETBERTKPICTFMNHSILVSGTGSNARCEVFNRESVSCQCHMTSFAVLMQ  
 VSRKNGELTLPKTLTYVALGVTLAALLTFPFLTLRLRNSONGIRNLTALGLA  
 QLVPLIGINQADLPACTVIALILHPLICTSPMALLEALHYRALTEKRDVNTGPMR  
 FLYLGMGVPAFTGLAVGLDEBEGNPFQWLSYDYLINFPAGVPAVAVMSLEYL  
 ILAABASCAOAGGFEKGVSGVGLDPSFVLLLSATWMLALSTNSDPLPHYLPAT  
 CNCTIGSPRIELSYVVLSEKVRKALKACRKSPPPALTTKSTLISYTCSPYADGR  
 LYOPTGDSAGSLHSTSRGSKQPSYIPLFLESALNPQGPGLADPESLFLQDQDQ  
 QHDPDTSDSLSLDDQSGSYASTHSDSEEBEERBEERBAFPBQGDLSLGGAB  
 RLPHSTPDGDPGPKAPWPDFTYAKESGNGAPERRLRENDALSRGSLGLPLP  
 GSSAOPHKGILLKKCLPTISEKSLRLPLBQCTGSSRSGSSASBESRGSPRRPPRO  
 SLQEQNGMPILAMSLKGTVEDSSGSFLFPNPLH"

BASE COUNT 2038 a 3353 c 2998 g 2142 t

## Alignment Scores:

Pred. No.: 0 Length: 10531  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-09-916-849a-3 (1-2923) x AF234887 (1-10531)

QY 1 MetAgsSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 Db 63 ATGCCGAGCCCGGCCACCGGCGTCCCTCCCAACCGCCGCGCTGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuGlyAAspGlnValGlyProCyAAspSerLeu 40  
 Db 123 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCyAlaProMetGlyTyrLeuCyAspSer 60  
 Db 183 GGGTCCAGGGGAGAGAGGCTCTTCCGGGGGCTGCGCCCAATGGGCTGGCTGCTCATTC 242  
 QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCyAAspAspAlaGlyTyrGluLeuThr 80  
 Db 243 TCAGGGCTCGAACCTGCTGCTTACACACGCGCTGAGGAGATGGGGCACTGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCyProGluSerGluHis 100  
 Db 303 GGCACCTGGTACCCACCAACATGGCTTGAAGGTTTGGTGTCCAGAAATCCAGGCGCAT 362  
 QY 101 IleProLeuProProAlaProGluGlyCyAspTyrSerCyArgGluLeuGlyTyrIleGly 120  
 Db 363 ATTCCCTTACCAACAGCTCTGAAGGCTGCCCTGAGGTGTGCGCTCTGGGCAATTGGA 422  
 QY 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGluHisProCyAspLeuVal 140  
 Db 423 GGCACCTTTCCCAACAGGGCAAGCTCACACTGCCGAGGAGACACCCGCTTAAAGGCT 482  
 QY 141 ProArgLeuArgCyGluSerCyAspLeuValGlnAlaProGluLeuValGlyGlu 160  
 Db 483 CCAAGGCTCAAGTGCAGTCTGCAAGCTGGGCAAGGCCCCCGGGCTCAAGGCAAGGAGA 542  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgGlyArgAsnValAsnThrAlaProGln 180  
 Db 543 AGGTACACAGAGAGTCCCTGGGTGGGGTGGGAAAGGAAATGTAAATACAGCCCCCAG 602

QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAAspGlnProAlaGlyTyrProVal 200  
 Db 603 TTCAGCCCCCAGCTTACAGGCCACATGTCGAGGAGAACACGACGAGGCACCCCCTGT 662  
 QY 201 AlaSerLeuValGalaIleAspProAspGluGlyValGlyValArgLeuGluTyrThrMet 220  
 Db 663 GCATCCCTGAGAGGCCATGACCCCGACAGAGGTGAGGAGGTGACCTGAGATACCAATG 722  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 723 GATGCCCTCTTGTGATNAGCCGCTCCACCAAGTCTTCTCCCTGAGACCACTGCTGTGCA 782  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrTyrSerThrHisValPheArgValThr 260  
 Db 783 GTAACCAACGCGGAGGAGCTGATGCTGAGCAAGAGACCAACCGTCTTCAGGGTCAAG 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 843 GCGCAGACACAGGATCCCGACAGAGTCCCTGAGCTACCTCACATCTTGATCT 902  
 QY 281 AspThrAsnAspHisAspProValPheGluGluGluGlyTyrGlySerLeuArgGlu 300  
 Db 903 GACACCAATGACATGACCTCTGTTCAGACGACGAGATGACAGAGAGCTCCAGGAG 962  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 963 AACCTGAGGTTGGCTATGAGGTGCTCACTGTCAAGGCGCACGAGTGTGATGCCCTCC 1022  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyGlySerProSerGluValPhe 340  
 Db 1023 AATGCAATATTCGTATCCGCTGCTGAGAGGCTGAGGAGGCTCCCTGCAATCTTT 1082  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 1083 GAGATCGACCTCCCTGCTGAGGTATCCGAACCCGTGCTGTGATCCGGAGAGGCT 1142  
 QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1143 GATCTCTTACAGCTGACGCTGATGAGGCAAGTACCAAGGCTCGGAGCCCGGCTCTGAGT 1202  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
 Db 1203 ACCAGGCGGCTGTTTCTCTTGTGTGAGATGACATGATGATGATGATGATGATGAT 1262  
 QY 401 GluIleArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db 1263 GAGAGGCTTATGTGTGCTCAAGTGAAGGAGATGTGACTCCAGGGGCCCAAGTACTCGCA 1322  
 QY 421 ValThrAlaSerAspArgAspGlySerAsnAlaValHisTyrSerIleMetSer 440  
 Db 1323 GTCAAGGCTCGATCGAGACAGAGGAGCAATGCCGTGTCATATGATCATGAGT 1382  
 QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 Db 1383 GGCATATGCTCGGAGACAGTTTATCTGATGATCCCAAGCTGAGCTGAGATGTGTGAGC 1442  
 QY 461 ProLeuAspTyrGluThrThrLeuGlyTyrThrLeuArgValAlaGlnAlaAspGlyGly 480  
 Db 1443 CCTCTGATCTATGAGACGACCAAGAGATGACCTTACCGGTGCCAGACAGATGTGTGC 1502  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1503 GGTCCCCCACTCTATATGTCTCTGGGCTGTGTGACAGTACAGTCTCTGATATACAGAC 1562  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 1563 AATGCCCATCTTCTGTGACGACCCCTTTCCAGGCTTATGCTCTGAGAGGCTCCCTTA 1622  
 QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db 1623 GGCTACCTGGTTCATCATGTCACAGCTATGAGCTGATGTGTGACATAGCCGCTG 1682  
 QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560

```

Db      1683 GAATACCGCTCTGCGGGGAGCATGACTCCCTTCACATCAACATGGCAGGCG 1742
Qy      561 ThrIleSerValAlaIleuAspArgGluValAspPheTyrSerPheGlyVal 560
Db      1743 TGGATCTCTGTGGCTGTGTAACGAGCCGGAGAGATGATTTCTTACACCTTGGGGTA 1802
Qy      581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      1803 GAAGCTCGAGACCATGGCACTCCAGCACTCACTGCCCTGGCCAGTGCAGCTGAGCTGTC 1862
Qy      601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      1863 CTGGAGTCTCAACACACAACTCAACCTTTACCAACAGATGACAGTGGCGGCTCAAT 1922
Qy      621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      1923 GAGGATGCACTGTGGGACACGCGTGTGACGCTGTGACCTGTGACCGTGAATGCTCAT 1982
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1983 AGTGATATCACTTACCAAGATCAACAGTGGCAATATCTGAAACCGCTTCTTCATCAACAGC 2042
Qy      661 GluSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluValGln 680
Db      2043 CAAGTGTGTGTGGCTGTATCTTCCTTGCCTGCACTGCACTCAAACTTGAAGCGGAG 2102
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 700
Db      2103 TATGTGTGGCTGTATCCGCTCCGATGGCACTCGGAGAGACCGGACAGATTTGGTG 2162
Qy      701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerHisTyrThrValAsn 720
Db      2163 AATGTACCGACCGCAACCCCATCGTCTGTCTTCAAGAGCTCCCATCAAGTGAAT 2222
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db      2223 GTTAATGAGACCGGCGGAGGACACACGCTGTGTGATCAGCGGCAAGATGAGAGAC 2282
Qy      741 ThrGlyGluAsnAlaArgGlyLeuThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2283 ACAGGTGAGATGCGCGCATCACTTACTTATGAGAGACAGCATCCCGGATTCGCGATC 2342
Qy      761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2343 GATGACAGACCGGGGGCTGCACCAACGAGCTGAGCTGACAGAACCAAGTGTCT 2402
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db      2403 TACACCTGCGCATTTACTGCTCGGAGCAATGGCAATCCCAAGAGTCCGACACCTAC 2462
Qy      801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSerTyr 820
Db      2463 CTGGAGATCTCTGTGAACGAGATGATGACATGCCCTTCACTGCTGCGAGACTCTTAC 2522
Qy      821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2523 CAGGGCAGTGTCTATGAGATGTGCCACCTTCACTAGCGCTCGAGATCTCAGCACT 2582
Qy      841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db      2583 GATCGTATTTCTGACTTAATGCGAGGGATCTTCAACCTTCAAGAGGCGAGAGTGA 2642
Qy      861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880
Db      2643 GACGGTACTTTATTTGTTGATGCTCACTGAGCATCTGCGAAGCGTCAACGAGGCTGAT 2702
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaAspIleGlyMetProPro 900
Db      2703 CGAAGAGACGTCGCCAGTATGTCTTGTGGCATATCAGTGAACAGGGGATGCCCCCA 2762
Qy      901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920

```

---

```

Db      2763 GCCCGACACCTATGAGAAATGACAGTCACTGTGTGATGTGAAATGACAAATCCCTGTCTC 2822
Qy      921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2823 TTGAGACAGATATGATTTGATGTGTTGTGAAAGAGACAGCCCATTTGGGCTTACCGGTC 2882
Qy      941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      2883 GCCCGGTCAAGACCATGACCCCGATGAGGACCAATAGCCCAATGATTTATGACAGATT 2942
Qy      961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla 980
Db      2943 GTGAGGGGCAATCTCGAGGGTCTTCAAGCTGACATCTTCTCCGGGAGCTGACAGCC 3002
Qy      981 LeuValAspLeuAspTyrGluAspArgProGluTyrValIleValIleGlnAlaThrSer 1000
Db      3003 CTGTGATCTTGAATGACAGAGACCGGCTGATGATGCTCTGTATCCAGGCACTGCA 3062
Qy      1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020
Db      3063 GCTCCTCTGTGAGCGGGCTTACAGTCACTGCGCTTCTTGAACGCAATGACAAACCA 3122
Qy      1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040
Db      3123 CCAAGTGTGGGCACTTGTGATCTTTCACCAACTATGTCACCAATGCTTCAAGCAAGC 3182
Qy      1041 PheProGlyGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3183 TTCCCTGGGGGTGCCATTTGGCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3242
Qy      1061 ThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      3243 ACTTACAGCTTTGAGGGGGAATGAACTCAACCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3302
Qy      1081 GluLeuValLeuSerArgAlaLeuAspAsnAsnAspProLeuGluAlaIleMetSerVal 1100
Db      3303 GAGCTGAAGCTTAAGCGCGCATGAGCAACACCGGCTCTGTAGGCAATGAGCGG 3362
Qy      1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120
Db      3363 CTGTGTCTGAGAGGGGTACAGCGGTGACGCGCCAGTGGCGCTGTGTGACATCATC 3422
Qy      1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140
Db      3423 ACCGATGAGATCTCAACCCACAGCATCACTGCGCTGTGAGAGCAATGACCGAGCC 3482
Qy      1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrPro 1160
Db      3483 TTCTGTCTCAACATGTCTAGGCTCTTCACTCAGCGGATGAGCGGCAACGCTGGCCAGCGCA 3542
Qy      1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180
Db      3543 CGGAGACCACTGTGTGTCTTCAACGTAACGCGGACACGAGCGCCCGGGGGCCACATC 3602
Qy      1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyIleProProPheLeu 1200
Db      3603 CTCAAGGTAGGCTGTGCTGTGGGCGAGCGGCGGCGGCGGCGGCGGCGCTTCTCTG 3662
Qy      1201 ProSerGluAspLeuGlnIleArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220
Db      3663 CCTCTGAGGACCTGAGAGGCGCTTATCTTCAACCGAGCTGCTGACGCGCATCTG 3722
Qy      1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGlyAsnTyr 1240
Db      3723 GCAACAGCGGTGTGCTTTCAGCAACATCTGCTCGGAGAGCCCTGCGAGAACTAC 3782
Qy      1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260
Db      3783 ATGCGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3842
Qy      1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280
Db      3843 GTCTCTTCCGGCCATCAACCCGTCGAGAGGCTGCGCTGCGCTGCGCGCGCGCTTCTC 3902

```



QY 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysEtySerArgProCysGlyProHis 1300  
 DB ACCGGTGAAGTACTGACGAGACCGAGGTGAGCTCTGCTACTCGCGGCCCTGTGGCCCCAAC 3962  
 QY 1301 GlyArgCysArgSerArgGluGlyYrYrThrCysLeuCysArgAspGlyYrYrThrGly 1320  
 DB GGGCGCTGGCCGACGCGGCGAGGGCGGCTACACTGCTCTGTGTATAGTGTACAGGGGT 3963  
 QY 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysLysAsnGly 1340  
 DB GAGCACTGTGAGGTGAGTGTCTCCCTACGCGCCCTTGCAACCCCGGGTGTCTGCAAGATTGG 4023  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPhe 1360  
 DB GGCACCTGTGTCAACTGCTGTGGCGGGTTTCAAGTGGCATTTGCCATCTGGAGACTTC 4083  
 QY 1361 GlyLysProTyrCysGluValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 DB GAGAAAGCCCTACTGCGCAGGTGACACGCGAGCTTCCCGCCCATCTTCATCACCTTT 4143  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
 DB CGGGGCTGTGCGCAGGCTTTCCACTTCACTGGCCCTCTCTGTTCCCAAGAGAGCCG 4203  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyLysHisAspPheValAlaLeuGlu 1420  
 DB GACGGGTGTCTGTGTATGCAATGGCGCTTCAATGAGAGATGATGCTTGTGGCCCTCGAG 4263  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLysThrThrThrValSer 1440  
 DB GTGATTCAGAGAGAGGTTCAGCTCACCTTCTGTGCGGGAGTCAACACACGCGTGTCC 4323  
 QY 1441 ProPheValProGlyValSerAspGlyGlnThrPheIleThrValGlnLeuLysTyr 1460  
 DB CCATTCTGTCCCGGAGAGATCACTATGTGCCATGTGCAATGCGTGCAGTGAATCTAC 4383  
 QY 1461 AsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLysValAla 1480  
 DB AATAAGCCCACTGTGGGTGACAGAGGGCTCCCAAGGGCCCATCAGAGCAAGAGTGGCT 4443  
 QY 1481 ValIleValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB GTGGGTGACCGTGTGATGTGTGTGACACAGAGGTGGCTTGTGCTTCGAGTCTCTCGGGC 4503  
 QY 1501 AsnTyrSerCysAlaIleAsnGlyThrGlnGlyGlySerLysLysSerLeuAspLeuThr 1520  
 DB AACTACTCTGTGTCTGCCAGGGCACCCAGGGTGGCAAGAGTCTGTGAATCTGACG 4563  
 QY 1521 GlyProLeuLeuLeuGlyValProAspLeuProGlyLysPheProValArgMetArg 1540  
 DB GGGGCCCTCTACTAGGGCGGGGTGCTGACTGCCGAGAGCTTCCAGTCCGAATGGCG 4623  
 QY 1541 GlnPheValGlyCysMetLysArgAsnLeuGlnValAspSerArgHisIleAspMetLysAsp 1560  
 DB CAGTTCGTGGGTGTGACGTGGAACTGTGAGTGGACGCGGCACTATACATGTGTGAC 4683  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysValAsnValCysAspSer 1580  
 DB TTCATTGCCAACAATGGACCTGTGCTGTGCTGCTGCAAGAAACGTGTGTGACACG 4743  
 QY 1581 AsnThrCysHisAsnGlyLysThrCysValAsnGlnTyrAspAlaPheSerCysGlyCys 1600  
 DB AACACTGTGCACAAATGGGGGCACTTGTGTGAACAGTGTGGAGCGCGTTGAGTGGAGTGC 4803  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
 DB CCCCTGGGCTTGTGGGGGCAAGAGCTGTGCGCCAGAAATGTCCAAATCAGAGCATTCCTG 4863  
 QY 1621 GlySerSerLeuValAlaThrPheGlyLysSerLeuProIleSerGlnProTyrPheLeu 1640  
 DB GGCAGAGAGCTGTGTGGCTGTGGATGGCTCTGTGCTGCCATCTCCCAACCTGTGTACTTC 4923

QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 DB AGCTTCATGTTCCGACCGGCGCCAGGCGCAAGGTGTCTGTGTCCAGGCAATTCACAGGGGG 4983  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 DB CGCAGACCACTACCACTTACAGTACGAGAGGGCCAGGTATCTGACGCTGAGAGGCAACA 5043  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 DB GGGCTTCAGGCTCTCTCTCTCTGTGTGAGCCAGGCGGGCCAAATGACGTATCTGGCAC 5103  
 QY 1701 HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 DB CATGCACAGCTGTGCACCTGGAGCCAGCGGGGGCTGTGGCAATGCCATTCGTCTTCAT 5163  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLysLeuHisLeuSer 1740  
 DB TATGGGCAAGAGAGAGAGAGGCAACTGGGCCCCCGGCTGTGCATGTCTGCACCTGAGC 5223  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 DB AACATTAAGTGGCGGAGAACTGTGGCCAGCGCGGTGTGGCCGTGTGCTTGGGGC 5283  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 DB TGTTTGCAGGGGTGTGCGGTGAGCATACGCCAGAGGGGGTTTACAGCTGTGATCCAGC 5343  
 QY 1781 HisGlyGlyLysThrIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB CATGGGAGAGACATCAAGGTGAGCAGAGGCTGTACCTGTGACCTTGTGACCTTAAC 5403  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyrSerCysSerCysAsp 1820  
 DB CCGTCTCTGTACAGCTATTTGACGACAACTGTGGGCAAGCTATTTCTTGACCTGTAT 5463  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840  
 DB CCAGGTACTATGTGGCAACTGTACTATGTGTGTGACTGTGAACCGTGTGAGCAGCAG 5523  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 DB TCTGTGTATACCCGCAAGCCAGTGTGCCCATGTGCTTACTGTGAGATGTCCCCCAAT 5583  
 QY 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrThrPheGly 1880  
 DB TACCTTGGCCATACTGTGAGAACAGAGTTGACAGAGCTTGTCCCGTGGCTGTGGGGA 5643  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn 1900  
 DB CATCCCAATGTGGCCCATGTGCACTGTGATGTGCAAGAAAGCTTTGACCCAGACTGCAAC 5703  
 QY 1901 LysThrSerGlyGlyLysCysHisCysLysGlyLysAsnHisGlyTyrArgProProGlyLysProThr 1920  
 DB AAGACAAAGCGGAGGTGTGACTGTGACAGAGAAACCACTACCGGCCCCAGGAGGCCCAAC 5763  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyLysAsp 1940  
 DB TGCCCTTGTGTGACTGTATCCCAAGAGCTCTTGTTCAGAGTGTGTGACCTTGAGAT 5823  
 QY 1941 GlyLysCysProCysLysProGlyValIleGlyValArgGlnCysAspArgCysAspAsnPro 1960  
 DB GGCAGTGTCCATGACAGCCAGAGTGTGATCGGGGCTCAGTGTGACCGGTGTGACAAACCT 5883  
 QY 1961 PheAlaGlyValThrThrAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaIle 1980  
 DB TTTGTGTAGTGCACCAACAATGTGTGAGATGAAATTTATGACAGTGTGCCACAGAGCAAT 5943  
 QY 1981 GluAlaGlyIleTyrTyrProArgThrArgPheGlyLeuProAlaAlaIleProCysAspPro 2000  
 DB GAGGTGTGAGATCTGTGTGGCCCCGTATCCGCTTGGGCTGTGCTGTGCTCTCTGTGCC 6003  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGluHisArgGlyTyrPheLeuProPro 2020

Db 6063 AAGGCTCTTGGAGCTGCTGGCCAGCTGAGAGACAGAGGGGTGGCTCCCCCA 6122  
Qy 2021 AsnLeuPheAsnCyThrSerIleThrPheSerGluLeuIleuValPheAlaGluArgLeu 2040  
Db 6123 AACCTTCACTGACGACGCTGATCACTTCTCAAGAACTGAAGGGCTTCGCTGAGCGGCTA 6182  
Qy 2041 GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGluLeuAlaLeuLeuArg 2060  
Db 6183 CACGGGAATGATCAAGGCTAGACTCAGGGCGCTCCAGCAGCTAGCCCTGCTCTGGCG 6242  
Qy 2061 AsnAlaThrGlnIleSthraIagIYrPheGlySerAspValIleValAlaIYrGlnLeu 2080  
Db 6243 AAGCCACGAGACACAGCTGCTGATCTTGAGAGCAGCTCAAGGGGCTTCCAGCTG 6302  
Qy 2081 AlaThrArgLeuAlaIleIleGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6303 GCCACGGGCTGCTGGCCACGAGAGACCCACGGGGGCTTTGGGCTGTCTGCAACAG 6362  
Qy 2101 AspValIlePheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6363 GAGGTGACTTCACTGAGAACTGCTGGGTGGGAGCGCCCTCTGAGCAGCCCAAC 6422  
Qy 2121 LysArgIleThrGluLeuIleGlnGlnThrGluGlyGlyThrAlaThrPheLeuGlnHis 2140  
Db 6423 AAGCGGACCTGGAGCTGATCCAGCAGACAGAGGTGGCACCGCTGGCTGCCAGCAC 6482  
Qy 2141 TYGCluAlaIYrAlaSerAlaLeuAlaGlnAspMetArgIleSthrTYrLeuSerProPhe 2160  
Db 6483 TATGAGGCTTACGCCAGTGCCTGGCCCAAGACATGGGGACACCTTACCTAAGCCCTTC 6542  
Qy 2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspIleGlyAsnPheAla 2180  
Db 6543 ACCATCGTCAAGCCCAATGTCATCTCGTAGTGGCTTGGACAAAGGGAAGCTTGGCT 6602  
Qy 2181 GlnAlaIleLeuProArgTYrGluAlaLeuArgGlyGluGlnProProAspLeuGluThr 2200  
Db 6603 GGGGCGAGGCTGCGCCGTCAGAGGCTCGTGGGAGAGAGCCCGGAGCTTGGAGACA 6662  
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
Db 6663 ACGATCTTCTGCTGAGTGTGTCTTCAAGAGACGCCGCCGCTGTCAGGCCCGAGAGC 6722  
Qy 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGlnArgArgIleProGluLeu 2240  
Db 6723 CCGGAGAGGCGCCAGGACAGAGAGCTGACGAGGAGACGCGCACCGCGAGCTG 6782  
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTYrArgThrLeuIleGlyLeuLeuPro 2260  
Db 6783 AGCCAGGGTAGGCTGTGGCCAGGCTCATCTACCGCACCCCTGGCGGGCTAAGCTGCT 6842  
Qy 2261 HisAsnTYrAspProAspIleArgSerLeuArgValProIleAsnThr 2280  
Db 6843 CATTAATTAACCTTGAACAAGCGAGCTTGAAGTCCCAACGCCGATCAATACACA 6902  
Qy 2281 ProValIleSerIleSerValIleAspAspGluGluLeuLeuProArgAlaLeuAspIle 2300  
Db 6903 CCGGTGTGAGCATCAAGCTGATGATGATGAGAGCTTCTGGCCCGGGCCCTGAGCAAA 6962  
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleAspProIleVal 2320  
Db 6963 CCGGTCAAGGAGCTTCCGCTGCTGAGACAGAGAGCGAGACCAAGGCCATCTGTGTC 7022  
Qy 2321 PheTYrAsnHisSerIleLeuValSerGlyThrGlyGlyTYrPheSerAlaArgGly 2340  
Db 7023 TTTCTGGAACCATTCATCTGTGTGATGGGACAGGTGGCTGTGGCCAGAGGCTGGAA 7082  
Qy 2341 ValValIlePheArgAsnGluSerIleValSerGlyGlnAsnHisMetThrSerPheAla 2360  
Db 7083 GTCTCTTCCGCAATGAGAGCAGTCACTGCGCAAGGCAACCAATGAGAGGCTTGTCT 7142  
Qy 2261 ValLeuMetAspValIleSerArgArgGluAsnGlyGluIleLeuProLeuIleSthrLeuThr 2280

Db 7143 GTGCTCATGACGTTTCTCGCGGAGAAATGGGAGATCTGCCACTGAAGACTGACA 7202  
Qy 2381 TYrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu 2400  
Db 7203 TACGTGGCTCTAGAGTGTACCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7262  
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
Db 7263 TTGGCTATCTTGGCTGCAACCAACAGGCACTCCAGCTGAACCTGACAGCTCCCTGGGCG 7322  
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCyThr 2440  
Db 7323 CTGGCTCAGGTGTCTTCTCTGGGAATCAACAGGCTGACCTCTTGTGCTGGACA 7382  
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTYrLeuCyThrPheSerTYrAlaLeuLeuGln 2460  
Db 7383 GTCATTTGCATCTGTGACTCTGTATCTCTGTCACTTTTCTGGGCTGTGGAGG 7442  
Qy 2461 AlaLeuHisLeuTYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
Db 7443 GCTTGGACCTGTACCGGCACTCACTGAGGTGGCGGATGTCAACACCGGCCCAATGCGC 7502  
Qy 2481 PheTYrTYrMetLeuGlyTYrGlyValProAlaPheIleThrGlyLeuAlaGlyLeu 2500  
Db 7503 TTTCTACTACATCTGGGCTGGGGCTGTGCTTCAATCAAGGGCTAACCGTGGGCTG 7562  
Qy 2501 AspProGluGlyTYrGlyAsnProAspPheCyThrPheSerIleTYrAspThrLeuIle 2520  
Db 7563 GACCCGAGGGCTTAGGGAACTTGAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7622  
Qy 2521 TTPSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTYrIleLeu 2540  
Db 7623 TGGAGTTTGTGCTGGCCGGGTGGCTTTGGCTCGATGAGAGTCTTCTGTCATCTCG 7682  
Qy 2541 AlaAlaArgAlaSerCyAlaAlaGlnArgGlnIlePheGluValArgValProAlaSer 2560  
Db 7683 GGGGCCGGGCTCTGTGCTCCAGGGCCAGGGCTTGAAGAAAGTCTGTCTG 7742  
Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTYrLeuLeuAlaLeu 2580  
Db 7743 GGCCTGAGCCCTCTTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7802  
Qy 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTYrLeuPheAlaThrCyAsnCyIle 2600  
Db 7803 CTCTGTCTCAACAGGACACCTCTCTTCACTACCTCTTGTCACTGCAATTCATC 7862  
Qy 2601 GlnGlyProPheIlePheLeuSerTYrValIleLeuSerIleGlyValArgIleAlaLeu 2620  
Db 7863 CAGGGCCCTTATCTTCTCTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7922  
Qy 2621 LysLeuAlaCySerArgIleAspProAspProAlaLeuThrThrIleSerThrLeu 2640  
Db 7923 AAGCTTGCTTGCAGCGGAGCCAGCCCTGACCTGCTGACCAACCAAGTCCACCTG 7982  
Qy 2641 ThrSerSerTYrAsnCyEProSerProTYrAlaAspGlyArgLeuTYrGlnProTYrGly 2660  
Db 7983 ACTCTCTTCAACCTGCCCCAGCCCTTACCGAGATGGCGGCTGTACCAAGCTTACGGA 8042  
Qy 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTYrIle 2680  
Db 8043 GACTGGCGGCTCTCTGACAGCACTGCTGCGGCAAGATCAAGCCAGCTACATC 8102  
Qy 2681 ProPheLeuLeuArgGluIleSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db 8103 CCTTCTTGTCTGAGGAGAGTCCCACTGAACCTTGGCCAGAGGGCCCTTGGCTGGGG 8162  
Qy 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8163 GATTCAGGAGGCTGTTCTTGAGAGGTCAAGACAGACAGATCTTACACAGGACTCC 8222  
Qy 2721 AspSerAspLeuSerLeuGluIleAspAspGlnSerGlySerTYrAlaSerThrHisSerSer 2740  
Db 8223 GACAGTGACTGTCTTAAAGACAGACAGAGTGGCTCTTATGGCTTACCACTCATCA 8282





QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 DB 1081 GAATTCACAGCGTGAAGGAGCAAGTGAACAGGGTGGAGCCCGGATCTCCGAGT 1140  
 QY 381 ThrThrAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 DB 1141 ACCAAGCCCGCTGTTTCTTCTGTGTGAAGATGACATATATATGCCCCCAGTTTAAAT 1200  
 QY 401 GluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 DB 1201 GAGAGGCGCTATGTGTGCAGGTGAGAGAGATGTGCTCCAGGGGCCCAAGTACCCGA 1260  
 QY 421 ValThrAlaSerAspAspAspLysGlySerAspAlaValHisTyrSerIleMetSer 440  
 DB 1261 GTCAAGCCCTCGATCGACAGCAAGGAGCAATCCCTGGTGCATATACATCATAGAT 1320  
 QY 441 GlyAspAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 DB 1321 GGCAATGCTCGGGGACAGTTTATCTGATGCCCAGACTGAGCTGAGATGTGTAGC 1380  
 QY 461 ProLeuAspTyrGlnThrThrLysGlnTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
 DB 1381 CCTCTTACATAGACGACCAAGAGATGACCTCTACGGGTGCGAGCAAGAGATGTGAC 1440  
 QY 481 ArgProProLeuSerAspValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500  
 DB 1441 CGTCCCACTCTATATGTCTCGGCTGTGTGACATGACATGCTCGATATCAAGAC 1500  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 DB 1501 AATGCCCCCACTTCCTGTGACGACCCCTTCCAGGCTACTGCTCGAGAGACGTCCTTCA 1560  
 QY 521 GlnTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeu 540  
 DB 1561 GGCTACTGTGTCTCCATGTCAGAGCTATGACGCTATGTGTGCAATATGCCGCTG 1620  
 QY 541 GlnTyrArgLeuAlaGlyValGlnHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 DB 1621 GAATACGCGCTGTGCTGGGGTGGGACATGACTTCCCTTCCACATCAACATATGGACAGGC 1680  
 QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
 DB 1681 TGGATCTCTGTGTGTGCTGACCTGACCGGGAGGAACTGATTTCTTCAAGCTTGGGGTA 1740  
 QY 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerValSerValThrVal 600  
 DB 1741 GAAGCTGAGACCAAGGACTCCAGCACTCACTGCTCGGCAAGTCAAGGTGACTGTC 1800  
 QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlnTyrThrValArgLeuAsn 620  
 DB 1801 CTGGATGTCAACGACCAATCCAACTTTTACCCAAACAGAGTCAACAGTGGCGCTCAAT 1860  
 QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAlaAspArgAspAlaHis 640  
 DB 1861 GAGGATCAAGCTGTGGGCAACGAGGTGTGACGCTGTGACGCTGTGACCGTGAAGTCTCAT 1920  
 QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 DB 1921 AGGTCACTACCTACAGATACCAAGTGGCAATCTGAACACCGCTTCTCATACCAAGC 1980  
 QY 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGln 680  
 DB 1981 CAAGTGGTGTGGGCTGTGATCTTCCCTTCCGCACTGACATCAAACTTGAACGCGCAG 2040  
 QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 DB 2041 TATGTGTGTGTGTATCGCTCCGATGCGCACTGGGCAAGCAAGGCAACAAATGTGTGTG 2100  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 DB 2101 AATGTCAACGACGCAACCAACCAATGCTCTTCTTTCAGAGCTCCCACTATACAGTAAAT 2160

QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 DB 2161 GTTAATAGAGACCGGCGCGGCAAGCAACAGTGTGTGTATGACGGCCACGATAGAGAC 2220  
 QY 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 DB 2221 ACAGGTGAAGAACCCCGCATCACTTCAATGAGAGCAAGCATCCCCAGTTCCGCAATC 2280  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluAspTyrGluAspGlnValSer 780  
 DB 2281 GATGCAAGACCGGGGCTGTCAACCAAGGCTGAGCTGAGCTACGAAACCAAGTGTCT 2340  
 QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyr 800  
 DB 2341 TACACCTGTGGCATTAAGTCTGTGGGACATGSCATTCCCAAGATCGGACCAACCTTAC 2400  
 QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 DB 2401 CTGGAGATCTGTGTGAACGACGTGAATGACAAATGCCCTCAGCTTCTTGGAGACTTCTAC 2460  
 QY 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 DB 2461 CAGGCAAGTGTCTATAGAGATGCACTTCACTAGCTTCTGACAGATCTCAAGCACT 2520  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 DB 2521 GATCGTGAATTCGAACTTAATGGCAGGGTCTTCACTCACTTCCAAAGAGCGCAGTGA 2580  
 QY 861 AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 DB 2581 GACGGTGACTTTATGTTATGTTAGTCCACGTCAAGGCACTGTGCAACGTATGAGAGCTGAGT 2640  
 QY 881 ArgGluAspValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900  
 DB 2641 CAGAGAACCTGGCCCAAGTATGTTCTTGGGCAATATGACGTGACAAAGGAGATGCCCA 2700  
 QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProVal 920  
 DB 2701 GCCCGCACCTATGAGATGACATGCACTGTGTGATGATGATGATCAATCCCCCTGTC 2760  
 QY 921 PheGlnGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 DB 2761 TTTGAGCAGAGATGATGTTATGTTGTGTGAAGAAAGCCCAATGGGCTAGCCGTG 2820  
 QY 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 DB 2821 GCCCGGTCAACGACCACTGACCCCATGAGGCAACCAAGCCCAAGTATATGACAGATT 2880  
 QY 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980  
 DB 2881 GTGAGAGGCAACATCCCTGAGGTTCTTCAGCTGACATCTTCCCGGAGGCTGACAGCC 2940  
 QY 981 LeuValAspLeuAspTyrGluAspArgProGlnTyrThrValLeuValIleGlnAlaThrSer 1000  
 DB 2941 CTGGTAGACTTGAATACAGAGAGACGCGCTGAGTACGTCTGTGATCCAGGCAAGTGA 3000  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 DB 3001 GCTCTCTGTGTAGCGCGGCTACAGTCAAGTCCGCTCTTTCAGCCGCAATGACACCA 3060  
 QY 1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040  
 DB 3061 CCAAGTCTGGGCAACTTGTAGATCTCTTTCACCAATATATGTCACCAATGCTCAACACAC 3120  
 QY 1041 PheProGlyGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 DB 3121 TTCCTGGGGGTGCCATTTGCGAGTACCTGTGCCAAGACCTCGATATCTCAGATATGCTG 3180  
 QY 1061 ThrTyrSerPheGlnArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 DB 3181 ACTTACAGCTTTGAGGGGGAATGAACTCAAGCTGTGTCTGTCAATGCTTCAAGGGGT 3240  
 QY 1081 GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100

Db 3241 GAGCTGAAGCTTAAGCGCGGCACTGGACAACACCGGCTCTGGAGGCGCATCAAGAGTG 3300  
QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1120  
Db 3301 CTGGGTGCAGACGGCGGTACACAGGTTGACCGCCAGTGGCGCGCTGCTGGACATCATC 3360  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgGluLysAspSerProGluArg 1140  
Db 3361 ACCGAATGAGATGCTACCCACAGCATCACCTGGCGCTGGAGGACATGTCAACCGAGCGC 3420  
QY 1141 PheLeuSerProLeuGluLysPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
Db 3421 TTCCTGTACACACTGCTAGGCTCTTCATCAAGCGGTGGCCGCGACGCTGGCCAGCCA 3480  
QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
Db 3481 CCGGACACAGTGTGGTCTTCAAGTACAGCGGAGACCGGACCCCGCGGGGCGCACATC 3540  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPheLeu 1200  
Db 3541 CTCACAGTGAAGCTGTCTGGTGGCGCACGCCGACGGGCGCGGGGCGGCGCTTCCTG 3600  
QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuThrAlaIleSer 1220  
Db 3601 CCTCTGAGGACCTGGACGAGCGCTATACCTACCGCAGCTGTGACGGCCATCTCG 3660  
QY 1221 AlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240  
Db 3661 GACACACGGGTGGCTGCTTCGACGACACATCTGCTGGGAGCCCTCGAGAACTAC 3720  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProIleAlaSerSerSer 1260  
Db 3721 ATGCGGTGGGTGGTGGCTGCTGCTTCGACTCTCCGCGCGCTTCATCGCTCTCTCC 3780  
QY 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
Db 3781 GTGCTTTCCGGCCCATCCACCCCGTGGAGGCTGGCTGGCTGGCCCGCGGCTTC 3840  
QY 1281 ThrGlyAspTyrCysGluThrArgValAspLeuCysTyrSerArgProCysGlyProHis 1300  
Db 3841 ACGGAGTACTAGTCCGACGACCGAGGTGGACTCTGCTACTCGGGCGCTGGCCCCC 3900  
QY 1301 GlyArgCysArgSerArgGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
Db 3901 GGGCGTGGCGGACGCGGCGGCGGCTACACTGCTGTGTATGCTACAGGGGT 3960  
QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysAlaAsnGly 1340  
Db 3961 GAGCATGTGAGGTGAGTGTCTCCCTCAGCGGCTGGACCCCGGGGTGTGCAAGATGG 4020  
QY 1341 GlyThrCysValAsnLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
Db 4021 GGGACTGTGTCAACTGCTGGTGGCGGTTCATATGGCATTTGCCCATCTGGAGCTTC 4080  
QY 1361 GlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
Db 4081 GAGAAACCTTACTGCGCAGGTGACACGCGGAGCTTCCCGCCCATCTTCATCACTTT 4140  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
Db 4141 CGGGGCTGGCGGAGGTTTCACTTCACCTCGGCTCTCGTTTCCCAAGAGAGCGC 4200  
QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyLysHisAspPheValAlaLeuGlu 1420  
Db 4201 GACGGGTGCTGTGTATCAATGGCGCTTCATGAGAGATGACTTTGTGGCCCTCGAG 4260  
QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrThrValSer 1440  
Db 4261 GTGATCCAGAGAGGTTCAGCTCACTTCTGCGGGGAGTCAACACACAGGTGTCC 4320  
QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuLysTyrTyr 1460

Db 4321 CCATTCGTGCCCGGAGAGTCAGTATGCCAGTGGCATACGGTGCAGCTGAATATCTAC 4380  
QY 1461 AsnLysProLeuLeuGlyGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnLysValAla 1480  
Db 4381 AATTAAGCCATCTTGGGTGTAGACAGGGCTTCCACAGGGCCCATCAAGACAGAAAGTGGCT 4440  
QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4441 GTGTGTACCGGTGATGGCTGTGACACAGGAGTGGCTTGCCTTGCGATCTGTCCGTGGC 4500  
QY 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLysSerLeuAspLeuThr 1520  
Db 4501 AACTACTCTGTGTGTGCCACAGGACCCAGGGGTGGACAGCAAGATCTTGATTCGACG 4560  
QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluLysPheProValArgMetArg 1540  
Db 4561 GGGCCCTCTCTAGAGCGGGTGTGACTGTCCCGAGAGCTTCCAGTCCGAATGGC 4620  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
Db 4621 CAGTTCGTGGCTGTGATGGGAACCTGCAGGTGACAGCCGCGCATATGACATGGCTGAC 4680  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
Db 4681 TTCATTTGCCAACAATGGACCGGTGCTGGCTGCTCCCTGCACAAAGAACCTGTGTACAGC 4740  
QY 1581 AsnThrCysHisAsnGlyGlyTyrCysValAsnGlnTyrAspAlaPheSerCysGluCys 1600  
Db 4741 AACATTCGCCAACAATGGGGGCACTTGCCTGAAACAGGTGGAGCGCTTCACTGCGAGTCC 4800  
QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisAspLeu 1620  
Db 4801 CCCCTGGGCTTTGGGGGCAAGAGCTGCCGCCAGAAATGGCAATCCACAGCATTCCTG 4860  
QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640  
Db 4861 GGCACAGCTGTGTGCTGTGATGCTCTGCTGCTGCCATCTCCCAACCTGTGTACCTC 4920  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCTTCATGTTCCGACCGCGGCGGACGAGCTGTCTGCTGCAAGCATCCAGGGG 4980  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGluGlyThr 1680  
Db 4981 CGCACACCATCACTCTACAGCTACGAGAGGGCCACGTGATGCTGAGGTGGAGGACACA 5040  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
Db 5041 GGGCTTCAGGCTCTCTCTCTCGTGTGAGCCAGGCGCGGCAATGACGTGATCGGAC 5100  
QY 1701 HisAlaGlnLeuAlaLeuGlyValAserGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
Db 5101 CATGCACAGCTGGCACTGGAGGCCAGCGGGGGCCTGGCCATGCCATCTGTCTTCAT 5160  
QY 1721 TyrGlyGlnGlnAspGlyGlyLysLeuLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
Db 5161 TATGGCGACAGAGGACAGAGGCAACTGGGCGCCCGGCTGCATGTCTGCACCTGAGC 5220  
QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
Db 5221 AACATTAACGTGGGGAATACCTGGGCAAGCCGGCGGTGTGGCCCGTGTGGGGC 5280  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
Db 5281 TGTTTGAGGGGTGTGGGTGAGATACGCGCAAGAGGGGTTAACAGCTGGATCCAGC 5340  
QY 1781 HisGlyGlySerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGACATCACTAAGGTGAGCAAGCTGTACCTGTGACCTTGTGACCTCAAC 5400  
QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAsp 1820  
Db 5401 CGGTCTCTGCTTAACGCTATTGGACGACACATGGGACAGCTATTCTGCGAGCTGTAT 5460

QY 1821 ProGlyTyrTyrGlyAspAsnGlyThrAsnValCysAspLeuAsnProGlyLysGln 1840  
Db 5461 CCAAGTTACTATGTTGACCACTGTAATGTTGTTGACCTGAAACCGGTGTGAGCAACG 5520  
QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGlyCysProProAsn 1860  
Db 5521 TCTGTGTGTATCCCGCAAGCCCAAGTGTGCCCCCATGAGCTATACCTGCGAGTGTCCCAAAAT 5580  
QY 1861 TyrLeuGlyProTyrCysGlyLysThrArgLysAspGlyProCysProArgLysTyrPrtPrtGly 1880  
Db 5581 TACTTGTGAGCCATCTGTGAGAACCGAGATTGACAGCTTGTCCCGGTGCTGTGAGGA 5640  
QY 1881 HisProThrCysGlyProCysAsnGlyAspValSerLysGlyPheAspProAspCysAsn 1900  
Db 5641 CATCCCAACATGTGGCCCATGCACTGTGATGTCAAGCAAAAGCTTTGACCCAGACTGCAAC 5700  
QY 1901 LysThrSerGlyGlyCysHisGlyLysGlyLysAsnHisGlyTyrArgProProGlySerProThr 1920  
Db 5701 AAGACAAGCCGCGAGTGTCCATGCAAGAGAACCACTAACCGGCCCCAGAGCAAGCCACCC 5760  
QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyLysAsp 1940  
Db 5761 TGCCTCTGTGTGATGCTGACCCCAAGGCTCTGTGCAAGATCTGTGACCTGTGAGAT 5820  
QY 1941 GlyGlnCysProCysLysProGlyValLysGlyArgGlnCysAspArgCysAspAsnPro 1960  
Db 5821 GGCAGATGTCATGACCAAGCGAGGTGTCATGGGGGTGATGATGATGATGATGATGATGATGAT 5880  
QY 1961 PheAlaGlyValThrThrAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaLys 1980  
Db 5881 TTGTGTGAGGTGACCAACCAATGTGCTGAGATGATGATGATGATGATGATGATGATGATGAT 5940  
QY 1981 GluAlaGlyLysTyrPrtProArgThrArgPheGlyLeuProAlaAlaLysProCysPro 2000  
Db 5941 GAGGCTGGAGATCTGTGTGAGCCCGTACCGGCTGTGAGGTGTGTGTGTGTGTGTGTGTGTGT 6000  
QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlyLysAspGlyTyrPrtLeuProPro 2020  
Db 6001 AAGGCTCTTTGGAGCTGTGTGTGCGCACTGTGATGAGCAAGGGGGGTGTGTGTGTGTGTGTGT 6060  
QY 2021 AsnLeuPheAsnGlyThrSerLysTyrPheSerGlyLeuLysGlyPheAlaGlyLysLeu 2040  
Db 6061 AACCTCTTCAACTGACGTCATCACTTCTCAGAACTGAAAGGGCTTGTGTGTGTGTGTGTGTGT 6120  
QY 2041 GlnArgAsnGlyLysSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
Db 6121 CACCGGAATGATGACAGGCTTGACTCAAGGCGCTCCAGCAAGCTTACGCTGTGTGTGTGTGTGT 6180  
QY 2061 AsnAlaThrGlnHisGlyThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
Db 6181 AAGCGCACGAGCAACAGCTGTGCTACTTGGCAGAGCAAGTCAAGGTGGCTTACAGAGCTG 6240  
QY 2081 AlaThrArgLeuLeuAlaHisGlyLysThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6241 GCCACCGGCTGTGTGTGCGCCAGAGACCCACAGCGGGCTTGTGTGTGTGTGTGTGTGTGTGT 6300  
QY 2101 AspValHisPheThrGlnLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6301 GAGCTGCACTTCACTGAGATCTGT 6360  
QY 2121 LysArgHisGlyPrtGlyLeuLysGlnGlnThrGlyGlyTyrAlaTyrPrtLeuLeuGlnHis 2140  
Db 6361 AAGCGGCACTGGAGCTGATCCAGACAGAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6420  
QY 2141 TyrGlnAlaLysTyrAlaSerAlaLeuAlaGlnAsnMetArgHisGlyTyrLeuSerProPhe 2160  
Db 6421 TATGAGGCTTACCGCAAGT 6480  
QY 2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspLysGlyAsnPheAla 2180  
Db 6481 ACCATGTGTACGCGCAACATTTGTATCTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6540

QY 2181 GlyAlaLysLeuProArgTyrGlyLysLeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
Db 6541 GGGGGCCAACTGTCCCGGTACAGAGGCGGTGTGGGAGCAAGCCCGGACCTGTGAGACA 6600  
QY 2201 ThrValIleLeuProGlyLysThrValPheArgGlyLysThrProValValArgProAlaGly 2220  
Db 6601 AAGATCATTTCTGCTGAGT 6660  
QY 2221 ProGlyGlyAlaGlyLysLeuProGlyLysLeuAlaArgArgGlnArgArgHisAspProGlyLeu 2240  
Db 6661 CCGGAGAGGCGCCAGAGCCAGAGAGCTGTGCAAGCGGACAGAGGAGCAAGCCGAGAGCTGT 6720  
QY 2241 SerGlnGlyLysAlaValAlaSerValIleLysTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db 6721 AGCCAGAGGAGGCTGT 6780  
QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleLysAsnThr 2280  
Db 6781 CATTAATATGACCTGTGACAAAGCGAGCTTGAGAGATCCCAAGCGCCGATCATCAACACA 6840  
QY 2281 ProValIleSerIleSerValHisAspAspGlyLysLeuLeuProArgAlaLeuAspLys 2300  
Db 6841 CCGGTGTGAGCATGACGCTCATATGATGAGAGCTTCTGTCCCGGCGCTGTGAGACA 6900  
QY 2301 ProValThrValGlnPheArgLeuGlnGlnThrGlnGlnArgThrLysProIleCysVal 2320  
Db 6901 CCGGTACAGGTGCAATTCGCTGT 6960  
QY 2321 PheTyrAsnHisSerIleLeuValSerGlyTyrThrGlyGlyTyrPrtSerAlaArgGlyCysGly 2340  
Db 6961 TTCTGTGAAACATTTCAATCTGT 7020  
QY 2341 ValIlePheArgAsnGlyLysSerHisValSerCysGlnCysAsnHisGlyMetThrSerPheAla 2360  
Db 7021 GTGTCTTCCGAAAGAGAGGCAAGCTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7080  
QY 2361 ValLeuMetAspValSerArgArgLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2380  
Db 7081 GTGTCTATGAGCTTTCTGT 7140  
QY 2381 TyrValAlaLeuGlyValThrThrAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 2400  
Db 7141 TACGT 7200  
QY 2401 LeuArgIleLeuArgSerAsnGlnHisGlyLysArgArgAspLeuThrAlaAlaLeuGly 2420  
Db 7201 TTGCGATCTGT 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyLysLeuGlnAlaAspLeuProPheAlaCysThr 2440  
Db 7261 CTGGCTCAGCTGT 7320  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGln 2460  
Db 7321 GTCATTTGCCATCTGT 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValAlaSerAspValAsnThrGlyProMetArg 2480  
Db 7381 GCTTGTGACCTTACCGGCACTCACTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db 7441 TTCTACTATCTGT 7500  
QY 2501 AspProGlyLysTyrGlyLysAspProAspPheCysTyrPrtLeuSerIleTyrAspThrLeuIle 2520  
Db 7501 GACCCGAGAGGCTTACGGGAACTCTGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7560  
QY 2521 TyrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
Db 7561 TGGAGTTTGT 7620  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLysGlyProValSer 2560

```

Db      7621 GCGGCGCGGCTCTCTGTCCTCCAGCGGCGGCGCTTTGAAGAAAGGCTCTCTCG
Oy      2561 G1yleuGlnProSerPheAlaValleuLeuLeuSerAlaThrTrpLeuAlaIeu
Db      7681 GGCCTGAGGCGCTCTCTGCGCGTCTCTCGTGTGAGGCCACGTGCTGTGCACTG
Oy      2581 LeuSerValAsnSerAspThrLeuPheHisTyrLeuPheAlaThrCysAsnCysIle
Db      7741 CTCTGTGCACAGGACACCTCTCTCTCACTACCTCTTGTGCTACCTGCAATTGCAATC
Oy      2601 G1ng1yProPhe11PheLeuSerTyrVal11LeuSerTyrGluValArgTyrAlaIeu
Db      7801 CAGGCGCCCTTATCTCTCTCTCTATGTGTGCTTAGCAAGAGTCCGGAACACTTC
Oy      2621 LysLeuAlaCysSerArgTyrProSerProAspProAlaLeuThrThrLysSerThrIeu
Db      7861 AAGCTTGCTGAGCGCGGAGGCCAGCCCTGACCTGCTCTCAACCAAGTCCACCTG
Oy      2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspG1yArgLeuTyrGlnProTyrG1y
Db      7921 ACCTGCTCTTACACAGTCCCGCCAGCCCTTACAGAGTGGCGGCTGTACCGCCCTACGGA
Oy      2661 AspSerAlaG1ySerLeuHisSerThrSerArgSerG1yLysSerGlnProSerTyrIle
Db      7981 GACTGGCGCGGCTCTCTGACAGCACCAAGTCCGCGGCGAAGTCAAGCCAGCTACATC
Oy      2681 ProPheLeuLeuArgTyrGlnuSerAlaLeuAsnProG1yG1ng1yProProG1yLeuG1y
Db      8041 CCTTCTTGTCTGAGGAGGAGTCCGCACTGAACTCTGCGCAAGGCCCTCTGCGCTGAGG
Oy      2701 AspProG1ySerLeuPheLeuGlnG1yG1nAspG1ng1nHisAspProAspThrAspSer
Db      8101 GATCCAGGACGCTCTTCTCTGGAAGTCAAGACCAAGCATGATCTTGAACGAGACTCC
Oy      2721 AspSerAlaPheSerLeuGlnAspAspGlnSerG1ySerTyrAlaSerThrHisSerSer
Db      8161 GACAGTGAACCTCTCTTGAAGACACCAAGTGTGCTCTTACGCTTACCACTACATCA
Oy      2741 AspSerG1ng1yGlnG1yGlnG1yGlnG1yGlnG1yGlnG1yGlnG1yGlnG1y
Db      8221 GACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
Oy      2761 TrpAspSerLeuLeuG1yProG1yAlaG1uArgLeuProLeuHisSerThrProLysAsp
Db      8281 TGGGATAGCTGCTGCGGCGCTGAGACAGAGACGCGCTTCAAGTCTCCCAAGAT
Oy      2781 G1yG1yProG1yProG1yLysAlaProTrpProG1yAspPheG1yThrThrAlaLysG1u
Db      8341 GGGGCGCCGAGGCGCTGAGGAGGCGCCCTGCGCAAGACTTGTGGACCAAGCAAAAGG
Oy      2801 SerSerG1yAsnG1yAlaProG1yGlnArgLeuArgG1yAsnG1yAspAlaLeuSerArg
Db      8401 AGTATGGGAAAGGCGGCGCCCTGAGAGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGG
Oy      2821 G1ng1ySerLeuG1yProLeuProG1ySerSerAlaGlnProHisLeuG1y 2837
Db      8461 GAGGCGCTCTGAGGCGCCCTTCAAGGCTCTTCTGCGCAGCTCACAAAGT 8511

```

```

RESULT 5
Db      D87469      8924 bp      mRNA      linear      PRI 06-OCT-2001
DEFINITION Human mRNA for KIAA0279 gene, partial cds.
ACCESSION D87469
VERSION D87469.1 GI:1665820
KEYWORDS KIAA0279.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y.,
AUTHORS Chara, O., Tanaka, A., Kotani, H., Miyajima, N. and Nomura, N.

```

TITLE Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain

JOURNAL DNA Res. 3 (5), 321-329 (1996)

MEDLINE 97191544

PIRMBED 9039502

REFERENCE 2 (bases 1 to 8924)

AUTHORS Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.

TITLE Direct Submission

JOURNAL Submitted (27-AUG-1996) Otsu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)

FEATURES

Location/Qualifiers

1..8924

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="HA6133"

/sex="male"

/issue\_type="brain"

/clone\_id="pSPORT 1"

1..8924

/gene="KIAA0279"

<1..7227

/gene="KIAA0279"

/note="Similar to D.melanogaster cadherin-related tumor suppressor"

/codon\_start=1

/protein\_id="PBA13407.1"

/db\_xref="GI:1665821"

BASE COUNT 1727 a 2846 c 2504 g 1847 t



## ORIGIN

## Alignment Scores:

Pred. No.: 0  
 Score: 2408.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 82.38%

DB: 9

Length: 8924  
 Matches: 2408  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-916-849A-3 (1-2923) x DB7469 (1-8924)

QY 516 GluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGly 535  
 DB 1 GAAGAGGTCCTCCCTTAAGCTACCTGGTCTCCATGTCACAGGCTTACACCGCTGATGCTGCT 60  
 QY 536 AspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIle 555  
 DB 61 GACAAATGCCCGCTGGAAATACCGCCCTTGCCTGGGGTGGACATGACTTCCCTTACCAATC 120  
 QY 556 AsnAsnGlyTyrGlyTyrPheSerValAlaAlaGlyLeuAspArgGlyGluValAspPhe 575  
 DB 121 AACAAATGGACACAGGCTGATCTCTGGTGGCTGAACTGGACCGGAGGAAGTTGATTTC 180  
 QY 576 TyrSerPheGlyValGlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSer 595  
 DB 181 TACAGCTTGGGGGTAGAAAGCTCAGACCAATGGCACTTCAGCACTCACTGCTCGGCCAAT 240  
 QY 596 ValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnProGlyTyr 615  
 DB 241 GTTAGGCTGACTCTCTGATGTCACAGCAACAACTTACCTTACCAACAGAGTTC 300  
 QY 616 ThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaVal 635  
 DB 301 ACAGTCCGGCTCAATAGAGATGACAGCTGTGGCAGCAGCGTGTGACGGTGTCACTGTG 360  
 QY 636 AspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArg 655  
 DB 361 GACCGGATGCTCATAGTGCATACCTTACCAATGACCAATGCTCAATATCTCAACACCGC 420  
 QY 656 PheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyr 675  
 DB 421 TTCTCATCAACGACCAAGCTGGTGGCTGGTATCCCTTCCCTGCACTGCACTTAC 480  
 QY 676 LysLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThr 695  
 DB 481 AAACCTTAGCGGCGATGATGTGGTGGCTGTACCGCTCCGATGCACTCGGCGAGCACG 540  
 QY 696 AlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSer 715  
 DB 541 GCAACAGATGTGGTGAATGTACCGACGCAACCAACCAATGCTCTGTCTTTCAGAGCTCC 600  
 QY 716 HisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSer 735  
 DB 601 CACTATACAGTGAATGTATATAGAGACCGCGCGGAGGACCAAGGTGGTGCATGACG 660  
 QY 736 AlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIle 755  
 DB 661 GCCACGATGAGGACACAGGTGGAATGCCCGATCACTTCAATGGGAGCACACATC 720  
 QY 756 ProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlyLeuAspTyr 775  
 DB 721 CCCCAAGTTCGGATCGATGCAAGACAGGGGGCTGTCAACCAAGGCTGAGCTGAGCTAC 780  
 QY 776 GluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLys 795  
 DB 781 GAAGACCAAGTGTCTTACACCTGGGCAATTACTGCTGGGACAAATGGCAATCCCGAAG 840  
 QY 796 SerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPhe 815  
 DB 841 TCCGACCAACCACTTACGAGATCTGGTGAACGACGCTGATGACAAATGCCCCCTCAGTTC 900  
 QY 816 LeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerValLeu 835

DB 901 CTGCAAGACTCTTACACAGGAGGTCTATAGAGATGGCCACCCCTTCACTAGCCTCTG 960  
 QY 836 GlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGln 855  
 DB 961 CAGATCTCAGGACCACTGATCGTATTTCTGACTTAAATGCAAGGGTCTTTCACCTTCCA 1020  
 QY 856 GlyIleAspAspGlyAspGlyAspPheIleValGlySerThrSerGlyTyrLeuArgThr 875  
 DB 1021 GGAGGCGACGATGGAGCGGTGACTTTATGTGTGAGTCCAGTCAGGACATCGTGGAAAG 1080  
 QY 876 LeuArgValLeuAspArgGluAsnValAlaGlnTyrValIleAspGlyAlaTyrAlaValAsp 895  
 DB 1081 CTACGAGGCTGATCGAAGAACGTGGCCAGATATGCTTGGCGGGCAATATGCAATGGAC 1140  
 QY 896 LysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAspValAsn 915  
 DB 1141 AAGGGAGATGCCCCAGCCCGCACCTTATGGAAGTACAGTCACTGTGTGGATTGAT 1200  
 QY 916 AspAsnProProValPheGlnGlnAspGluAspValPheValGlyGluAsnSerPro 935  
 DB 1201 GACAAATCCCGCTGCTTGTAGCAGGATGATTTGATGTGTGTGGAAGAACAGACCC 1260  
 QY 936 IleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlyTyrThrAsnAlaGln 955  
 DB 1261 ATTGGGCTAGCCGTGGCCCGGGTCAACAGCACTGACCCCGATGAAAGCAACCAATCCAG 1320  
 QY 956 IleMetTyrGlnIleValAlaGlyValAsnIleProGluValPheGlnLeuAspIlePheSer 975  
 DB 1321 ATTATGTACAGATTTGTGAGGGGCAACATCCCTAGAGCTTCCAGCTGGAATCTTCC 1380  
 QY 976 GlyIleLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGlyTyrValLeuVal 995  
 DB 1381 GGGAGAGCTGACAGCCCTGTGACTTACACGAGACCGGCTGATAGCTCTGTGCTC 1440  
 QY 996 IleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAsp 1015  
 DB 1441 ATCCAGGCGCAGTCAAGCTCTCTGTGAGCGGGCTACAGTCAAGTCCGCTCTTAC 1500  
 QY 1016 ArgAsnAspAsnProProValLeuGlyValAsnPheGlnIleLeuPheAsnAsnTyrValThr 1035  
 DB 1501 CGCAATGACAAACCAACCAAGCTGGGCAACTTGTAGATCTTTTCAACATATGTAC 1560  
 QY 1036 AsnArgSerSerSerPheProGlyGlyValAlaIleGlyArgValProAlaHisAspProAsp 1055  
 DB 1561 AATCCCTCAAGAGAGCTTCCCTGGGGGTGCATTTGCCAAGTACCTGCCATGACCTTAT 1620  
 QY 1056 IleSerAspSerLeuThrTyrSerPheGluArgGlyValAsnGlyLeuSerIleValLeu 1075  
 DB 1621 ATCTCAGATAGTCTACCTTACAGCTTGAAGGGGAATGAATCTAGCCTGTGCTCTC 1680  
 QY 1076 AsnAlaSerThrGlyGlyLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGlu 1095  
 DB 1681 AATGCTTCAACGGGTGAGCTGAGCTTAAAGCCGCGCATGACAAACACCGGCTCTGAG 1740  
 QY 1096 AlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGlnCysValLeu 1115  
 DB 1741 GCCATCAATGAGGTGCTGTGTGACAGCGCGTACACAGCTGACCGCCAGTGGCCGCTG 1800  
 QY 1116 ArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAsp 1135  
 DB 1801 CGTGTGACATATACCAAGATGATGCTCACCCACAGCATACCTGCGCTGGAGAGAC 1860  
 QY 1136 MetSerProGluArgPheLeuSerProLeuGlyLeuPheIleGlnAlaValAlaAla 1155  
 DB 1861 ATGTCAACCGAGCGCTTCTGTCAACCACTGAGCTCTTCACTCAAGCGGTGGCCGCGC 1920  
 QY 1156 ThrLeuAlaThrProProAspHisValValValPheAsnValGlnArgAspThrAspAla 1175  
 DB 1921 ACGTGGCCACCGCACCGGACCACTGTGTGTCTTCAACGTAACAGCGGAGCACCGACGCGC 1980  
 QY 1176 ProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProGlyProGlyGly 1195

Dh 1981 CCGGGGGCCACATCTCAACGTGAGCCTGTGGTGGGCCAAGCCGCCAGGGGCCCGGGGGC 2040  
Qy 1196 G1yProBProBheupProSerGluAspLeuGlnGluArgLeuTyrlLeuAmaArgSerLeu 1215  
Db 2041 GGGCCGCCCTTCCGCTCGAGGACCTGAGAGAGGCGCTATACCTCAACCGCAGGCTG 2100  
Qy 1216 LeuThrAlaIleSerAlaGlnArgValLeuProBheAspApaenIleCyBLeuArgGlu 1235  
Db 2101 CTAGCGGCACATCTCGGACAGCGCGTGTGCTCTTCCACAGCAACATCTCTGCGGGAG 2160  
Qy 1236 ProCysGluAsnTyrlMetArgCysValSerValLeuAspBheAspSerSerAlaProPhe 1255  
Db 2161 CCTTCGAGACCTACATGCGCTGCGTGGGTGCTGCTTCCATCTCTCGGCGCTTC 2220  
Qy 1256 IleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArg 1275  
Db 2221 ATGCGCTCCTCGCTCGGCTCTTCCGGCCATCCACCCGTCGAGGGCTGCGCTGCGC 2280  
Qy 1276 CysProBProGlyPheThrGlyAspTyrlCysGluThrGluValAspLeuCyBtyrSerArg 1295  
Db 2281 TGCCCGCCCGCTTCACGGGTGACTACTGCGAGACCAAGGTGGACCTCTCTACTGCGG 2340  
Qy 1296 ProCysGlyProHisGlyArgCysArgSerArgGlnGlyGlyTyrlThrCysLeuCyBArg 1315  
Db 2341 CCTGTGCCCCCAAGGGCGCTGCGGACGCGGAGGGCGCTACACTGCTCTGTCTG 2400  
Qy 1316 AspGlyTyrlThrGlyGlnHisCysGluValSerAlaArgSerGlyArgCysThrProGly 1335  
Db 2401 GATGGCTACACGGGTGAGCACTGTGAGGTAGTGTCTGCTCAAGCCGTTGCAACCCGGT 2460  
Qy 1336 ValCysLeuAsnGlyGlyTyrlThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCys 1355  
Db 2461 GTCTGCAAGATGAGGGGCACTGTGTCAACTGTGTTGGCGGGTTCAAGTCGCAATTGG 2520  
Qy 1356 ProSerGlyAspBheGlyLysProTyrlCysGlnValThrThrArgSerPheProAlaHis 1375  
Db 2521 CCACTGAGACCTTCAGAGAGCCCTACTGCGAGGTGACACAGCGACTTCCCGCCAC 2580  
Qy 1376 SerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPhe 1395  
Db 2581 TCTTATCATCCTTTGGCGGCTGCGGCAAGCTTTCACCTTCACTGCGCTCTGTCTT 2640  
Qy 1396 AlaThrLysGlyLysArgGlyLeuLeuLeuTyrlAsnGlyValArgPheAsnGlyLysHisAsp 1415  
Db 2641 GCCACAAAGAGCGCAGCGGTGTGCTGTGTACATAGGCGTTCAATGAGAGACATGAC 2700  
Qy 1416 PheValAlaLeuGlnValIleGlnGlnGlnValAlaGlnLeuThrPheSerAlaGlyGlyLeuSer 1435  
Db 2701 TTGTGACCTCGAGGTGATCCAGAGCAAGGTCCAGCTCACTTCTCTGCAAGGGAGTCA 2760  
Qy 1436 ThrThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrVal 1455  
Db 2761 ACCACCAAGGTGCTCCCATCTCGTGGCCGAGAGTCAAGTATGGCCAGTGCATACGGTG 2820  
Qy 1456 GlnLeuLysTyrlTyrlAsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSer 1475  
Db 2821 CAGCTAAATACTACATAGACCACTGTGGGTCAACAAGGGCTCCCAAGGGCCCATCA 2880  
Qy 1476 GluGlnLysValAlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPhe 1495  
Db 2881 GAGCAGAGAGGTGCTGTGTGACCGTGTGATGCTGTGACACAGAGATGGCTTGGCCTTC 2940  
Qy 1496 GlySerValLeuGlyAsnTyrlSerCysAlaAlaGlnGlyTyrlThrGlnGlyGlySerLysLys 1515  
Db 2941 GAGTCTGTCTGGGCAACTACTCTGTGTGCTGCCCAAGGCAACCAAGGTGGCAGCAAGAG 3000  
Qy 1516 SerLeuAspLeuThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPhe 1535  
Db 3001 TCTCTGATCTGACGGGGCCCTGTCTACTAGCGGGGTGCTGACCTGCGCCAGACACTTC 3060  
Qy 1536 ProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHis 1555  
Db 3061 CCACTCCGAATGGGAGATTCTGTGGCTGCAAGCGAACCCTGAGGTGACAGCCGGAGC 3120

Qy 1556 IleAspMetAlaAspPheIleAlaAsnArgGlyThrValProGlyCysProAlaLysLys 1575  
Db 3121 ATAGACATGGCTGACTTCACTTTCGCAACATGGCAGCCTGCTGGCTGCGCCAGAGAG 3180  
Qy 1576 AsnValCysAspSerAsnThrCysHisAsnGlyGlyTyrlThrCysValAsnGlnThrAspAla 1595  
Db 3181 AACGTGTGTGACAGCAACTTGGCCAAATGGGGGGCACTTGGCTGAACAGTGGAGCGG 3240  
Qy 1596 PheSerCysGlyCysProLeuGlyPheGlyGlyLysSerCysAlaGlnGlyMetAlaAsn 1615  
Db 3241 TTGAGCTGAGAGTCCCCCTGGAGCTTTGGAGCGAAGCTGGCCAGAGAAATGGCCAT 3300  
Qy 1616 ProGlnHisPheLeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSer 1635  
Db 3301 CCAAGACACTTCTGAGGACAGAGCCTGTGGCTGGCAAGGCTCTCTGCTGCCATCTCC 3360  
Qy 1636 GlnProTyrlLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGln 1655  
Db 3361 CAACCTGGTACTCAGCTCATGTTCGCAAGCGCAGAGCGGCTGCTGTCGAG 3420  
Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisIleMetLeu 1675  
Db 3421 GCCATCAAGAGGGGGGAGCAGCAATCACTTACAGCTACAGAGGGCCACGTATGTG 3480  
Qy 1676 SerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyValArgAlaAsn 1695  
Db 3481 AGCGTGAAGGGACAGAGGCTTACAGCTCTCTCTCCGTGGAAGCAGGCGGCGCAT 3540  
Qy 1696 AspGlyAspThrPheHisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAla 1715  
Db 3541 GACGGTGACTGGCAACATGCAAGCTGGGCACTGGAGCGCAGGGGGGCGCTGGCCATGCC 3600  
Qy 1716 IleLeuSerPheAspTyrlGlnGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
Db 3601 ATTCTGCTCTTGATTTATGGGAGAGAGAGAGAGAGGAGAACTGGAGCCCGCTGCAAT 3660  
Qy 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
Db 3661 GGTCTGACCTTACGACATACATACATGGGCGGAAATCTGGGCGCAGCCGCGTGTGGC 3720  
Qy 1756 ArgGlyPheArgGlyCysLeuGlnGlyValaArgValSerAspThrProGlnGlyValaAsn 1775  
Db 3721 CGTGGCTTTCCGGGCTGTGTGAGGGGTGCGGGTGAAGCAATCCAGAGGGGGTTAAC 3780  
Qy 1776 SerLeuAspProSerHisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
Db 3781 AGCTGTGAATCCAGCCATGGAGAGAGCAATCAAGTGAAGCAAGGCTGTAGCTGCTGAC 3840  
Qy 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrlCysSerAsnAspThrAspSerTyrl 1815  
Db 3841 CTTTGTGACTAAACCGGTCTCTGTCTAACAGCTATTGAGACAGACTGGGACAGCTAT 3900  
Qy 1816 SerCysSerCysAspProGlyTyrlTyrlGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
Db 3901 TCTTCAAGCTGTGATCCAGGTACTATGTGTGCAACTGATCAATATGTGTGTACCTGAC 3960  
Qy 1836 ProCysGlnHisGlnSerValCysThrArgLysProSerAlaProHisGlyTyrlThrCys 1855  
Db 3961 CCGTGTGACACAGTCTGTGTATCCCGCAAGGCCAGTGGCCCCCAAGGCTATACCTGC 4020  
Qy 1856 GluCysProProAsnTyrlLeuGlyProTyrlCysGlnThrArgIleAspGlnProCysPro 1875  
Db 4021 GAGTGTCCCCCAATTACTTGGGCAATACGTGTGACCAAGGATTGACCAAGCTTGTCCC 4080  
Qy 1876 ArgGlyTyrlThrPheGlyHisProThrCysGlyProCysAsnCysAspValaSerLysGlyPhe 1895  
Db 4081 CGTGTGCTGTGGGCAATCCCAATGTGCCCCCAATGAGCTGATGTGACCAAGGCTTT 4140  
Qy 1896 AspProAspCysAsnLysThrSerGlyGlyCysHisCysLysGluAsnHisGlyTyrlArgPro 1915  
Db 4141 GACCGAGACTGCAACAGCAAGCGGCGAGTGCCTGCAAGAGAAACATACCGGCCC 4200



Qy 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
Db 4201 CAGAGGAGCCCACTGCTCTTGTGTGATGCTGCAACCCCAAGGCTCTTGTCCAGATC 4260  
Qy 1936 CysAspProGlyLysArgGlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAsp 1955  
Db 4261 TGTGACCTTAGAGATGCGCAGTGTCCATGCAAGCCAGATGTCATCGGCGCTCAGTGTGAC 4320  
Qy 1956 ArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSer 1975  
Db 4321 CGCTGTACAACTTTTGTGAGGTCAACAATGCTGTGAAGTGAATATGAGACG 4380  
Qy 1976 CysProArgAlaIleGluValGlyIleTyrTyrProArgThrArgPheGlyLeuProAla 1995  
Db 4381 TGCCCAAGAGATGAGGCTGGGATCTGGTGGCCCGTACCGGCTTCGGGCTGGCT 4440  
Qy 1996 AlaAlaProCysProLysGlySerPheGlyThrAlaValArgHisCysAspGluHisArg 2015  
Db 4441 GCTGCTCCCTGTCCCAAGGCTCTTGTGGGATGCTGTGGCCACTGTGTAGACACAG 4500  
Qy 2016 GlyTyrLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGly 2035  
Db 4501 GGGTGGCTCCCCCAAACTTCTTCACTGCACTGTCATCACTTCTCAGAACTGAAGAGC 4560  
Qy 2036 PheAlaGluArgLeuGlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
Db 4561 TTGCGTAGAGGGCTACAGCGGAATGATCAGGCTTACCTCAGGGGCTCCAGACGCTA 4620  
Qy 2056 AlaLeuLeuLeuLysArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLys 2075  
Db 4621 GCCCTGCTCTGGGCAAGCGCAGCAGCAGCAGCAGCTGCTTACCTTCGACAGCGTCAAG 4680  
Qy 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGly 2095  
Db 4681 GTGGCTTCCAGCTGGCCACGCGGCTCTGGCCCAAGAGACACCGCGGGCTTTGGG 4740  
Qy 4741 CTGTCTGCCACACAGGAGCTGCACTTCACTGAGAAATCTGCTGGGGGAGCAGCGCTTC 4800  
Qy 2116 LeuAspThrAlaAsnLysArgHisTyrGlnLeuIleGlnGlnThrGlnGlnGlyThrAla 2135  
Db 4801 CTGACACAGCCCAACAGCGGCACTGGAGCTGATTCAGACAGAGGGGTGGACACGCC 4860  
Qy 2136 TyrLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetCysHisTThr 2155  
Db 4861 TGCGTGTCCAGCACTATGAGGCTTACGCACTGCTGCTGCGCCAGAACATGGCGACACC 4920  
Qy 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
Db 4921 TACCTAAGCCCTTCACTCATCTGTCAACCTTGTCACTCCGTAGTGGCTTGGAC 4980  
Qy 2176 LysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGlnAlaLeuLysArgGlyGlnGlnPro 2195  
Db 4981 AAAGGGAACTTGTGTGGGCGCAAGCTGCCCGCTACAGAGCCCTGTGGGGAGCAGGCC 5040  
Qy 2196 ProAspLeuGlnThrThrValIleLeuProGlyLysSerValPheArgGluThrProProVal 2215  
Db 5041 CCGGACCTTAGACAACTATCTTGTCTGAGTCTGTCTTCCAGAGAGAGCCCGCCGCTG 5100  
Qy 2216 ValArgProAlaGlyProGlyGlnGlnGlnGlnProGlnGlnLysArgArgGlnArg 2235  
Db 5101 GTACAGGCGCCAGAGCCCGGAGAGGCCCAAGAGCCAGAGAGCTGGCACGGCGACAGCGA 5160  
Qy 2236 ArgHisProGlyLeuSerGlnGlnGlnValAlaSerValIleIleTyrArgThrLeu 2255  
Db 5161 CGGAGCCCGAGGCTGAGCCAGGGGAGGCTGTGGCCAGCTCATCTACCCGACCTTG 5220  
Qy 2256 AlaGlyLeuLeuProHisAsnTyrAspProAspLysArgSerLeuLysArgValProLysArg 2275  
Db 5221 GCGGGGCTACTGCTCATTACTATGACCTGACAGAGCGAGCTTGAAGAGTCCCAAAAGC 5280  
Qy 2276 ProIleIleAsnThrProValValSerIleSerValHisAspAspGlnGlnLeuPro 2295

Db 5281 CCGATCATACACACCCGCTGGAGCATCAGGTCCATGATGATAGAGAGCTTCCGCC 5340  
Qy 2296 ArgAlaLeuAspLysProValThrValGlnPheArgLeuLeuGlnThrGlnLysArgThr 2315  
Db 5341 CGGGCCCTGAGCAAAACCCCTCAAGGTGAGTTCGCGCTCTGAGACAGAGAGCGGAGCC 5400  
Qy 2316 LysProIleCysValPheThrAsnHisSerIleLeuValSerGlyThrGlnGlyTyrPhe 2335  
Db 5401 AAGCCATCTGTCTTCTTGTGAACCATTCATACCTGTGTAGTGGACAGGTGGCTGGTGG 5460  
Qy 2336 AlaArgGlyCysGluValValPheArgAsnGluSerHisValSerCysGlnCysAsnHis 2355  
Db 5461 GCCAGAGGCTGTAGAGTGTCTTCCGACATGAGAGCAGTCAAGCTGCCAGTGAACAC 5520  
Qy 2356 MetThrSerPheAlaValLeuMetAspValSerArgArgGlnLysArgGlnIleLeuPro 2375  
Db 5521 ATGACGAGCTTGTGGTGTGTATGACGTTTCTCGCGGGAATGGGAGATCTTGCA 5580  
Qy 2376 LeuLysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhe 2395  
Db 5581 CTGAGACACTGACATACGTGGCTTGAAGTGCACCTTGGCTGCCCTTGTGCTACCTTC 5640  
Qy 2396 PhePheLeuThrLeuLeuArgGlyLeuArgSerAsnGlnHisGlyIleArgArgAsnLeu 2415  
Db 5641 TTCTTCTCACTCTTGGGATCTTGGCTGACGCTCCCAACACAGGCACTCCGACGTAC 5700  
Qy 2416 ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeu 2435  
Db 5701 ACAGCTGCTCTGGGCTGGCTCAGCTGCTTCTTCGGGAATCAACAGGCTGACCTTC 5760  
Qy 2436 PropheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSer 2455  
Db 5761 CTTTGTGCTGACAGTATTCATTCAGTCTGCTGCACTTCTGATCCTTGACCTTTTCC 5820  
Qy 2456 TyrAlaLeuLeuGlnAlaLeuHisLysTyrArgAlaLeuThrGluValArgAspValAsn 2475  
Db 5821 TGGGCTCTGAGAGGCTTGCACCTGTACCGGGCACTCAGAGGTTGGCCATGTCAAC 5880  
Qy 2476 ThrGlyProMetArgPheTyrTyrMetLeuGlnTyrGlnValProAlaPheIleThrGly 2495  
Db 5881 ACCGGCCCATCGCTTCACTACATGTGGGCTGGGGGTGGCTGCTCATTCACAGGG 5940  
Qy 2496 LeuAlaValAlaGlyLeuAspProGlnGlyTyrGlyAsnProAspPheCysTyrLeuSerIle 2515  
Db 5941 CTAGCGGTGGGCTGGACCCGAGGGCTACGGGAACCTTGACTTGTGCTGCTTCATC 6000  
Qy 2516 TyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSerVal 2535  
Db 6001 TATGACACGCTCATCTGAGGTTTGTCTGGCCGCTGGCTTGGCGTCTCATGAGTGTTC 6060  
Qy 2536 PheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLys 2555  
Db 6061 TTCTGTATCATCTGGGCGGCGCGGCTCTCTGTGCTCCACAGGGAGGAGCTTTGAGAG 6120  
Qy 2556 LysGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThr 2575  
Db 6121 AAAGTCTGTCTCGGGCTGGAGCCCTTCTTGTGCTCTCTGCTGTGAGCGGCAAG 6180  
Qy 2576 TyrLeuLeuAlaLeuLeuSerValLysSerAspThrLeuLeuPheHisTyrLeuPheAla 2595  
Db 6181 TGCGTGTGGCACTGCTCTGTCAACAGCAGCACCTCTCTTCCACTCACTCTTGTGCT 6240  
Qy 2596 ThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlu 2615  
Db 6241 ACTGCAATTTGATCCAGGGCCCTTCACTTCTCTCTATGTGTGCTTAGCAAGAG 6300  
Qy 2616 ValArgLysAlaLeuLysLeuAlaCysSerArgLysProSerProAspProAlaLeuThr 2635  
Db 6301 GTCCGGAAAGCACTCAAGCTTCTGTGACGCGCAAGCCAGGCCCTGACCTCTGACCC 6360  
Qy 2636 ThrLysSerThrLeuThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeu 2655

6361 ACCAAGTCCACCTCCTGCTCTACACACTGCCCCAGCCCTTACGACAGATGGGGCGCTG 6420  
QY 2656 TGTGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyYrYsser 2675  
Db 6421 TACCAAGCCCTACGAGAGACTCGGCGGCTCTGTGCAACAGACAGCTCGCTCGGGAAGAGT 6480  
QY 2676 GlnProSerTyrIleProPheLeuLeuArgGlyGlySerAlaLeuAspProGlyGlnGly 2695  
Db 6481 CAGCCCACTACATCCCCCTCTGCTGAGGAGGAGCTCGCACTAACCCCTGGCCAAAGG 6540  
QY 2696 ProProGlyLeuGlyAspProGlySerLeuPheLeuGlyGlyAspGlnGlnHisAsp 2715  
Db 6541 CCCCCGCTGGGGGATCCAGGACAGCTGTTCTCTGGAAGGTCAAGACCAAGACAGATGAT 6600  
QY 2716 ProAspThrAspSerAspSerAspLeuSerLeuGlyAspAspGlnSerGlySerTyrAla 2735  
Db 6601 CCTGACAGGAGACTCCGACAGTGAACCTGTCTTAGAAGACAGACAGATGGCTCTATGCGC 6660  
QY 2736 SerThrHisSerSerAspSerGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 2755  
Db 6661 TCTACCCCACTATCAGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6720  
QY 2756 ProGlyGlnGlnGlyTyrAspSerLeuLeuGlyProGlyAlaGlyArgLeuProLeuHis 2775  
Db 6721 CTTGAGAGAGAGAGGCTGGGATGAGCTGCTGGGGCTTGGAGACAGAGACTGCCCTTGAC 6780  
QY 2776 SerThrProTyrAspGlyGlyProGlyProGlyYrYsAlaProTyrProGlyAspPheGly 2795  
Db 6781 AGTACTCCCAAGATGAGGGGCCCAAGGGCTGGCAAGGCCCTGGCCAGAGACTTTGGG 6840  
QY 2796 ThrThrAlaYrYsGlySerSerGlyYrYsGlyYrYsAlaProGlyGlyGlyGlyGlyGly 2815  
Db 6841 ACCAAGCAAAAGAGAGATGAGGCAACGAGGGCCCTGAGAGAGCGCTCGGAGAGATGA 6900  
QY 2816 AspAlaLeuSerAspArgGlySerLeuGlyProLeuProGlyYrYsSerAlaGlnProHis 2835  
Db 6901 GATGCGCTGTCTCAGAGAGGGGTCTTACGGCCCCCTTCCAGGCTCTTCTGCGCAAGCTTAC 6960  
QY 2836 LysGlyIleLeuYrYsGlyCysLeuProThrIleSerGlyYrYsSerSerLeuLeuArg 2855  
Db 6961 AAAGGCATCTTAAAGAAAGATGTCTGCCACATCAGACGAGAGAGAGAGAGAGAGAGAG 7020  
QY 2856 LeuProLeuGlnGlnCysThrGlySerSerArgGlySerSerAlaSerGlyYrYsArg 2875  
Db 7021 CTCCCTCGAGAGATGACAGAGGCTTCTCCGGGGCTCTCCGCTAGAGAGAGAGAGAGAG 7080  
QY 2876 GlyGlyProProArgProProArgProProArgGlnSerLeuGlnGlnLeuAspGlyAla 2895  
Db 7081 GGGGGCCCCCTTCCCCGCCACCGCCCGGAGAGCTTCCAGAGACCTGAACGAGGCTC 7140  
QY 2896 MetProIleAlaMetSerIleYrYsAlaGlyThrValAspGlyAspSerSerGlySerGly 2915  
Db 7141 ATGCCCATCGCATGAGCATCAGGCAAGGCAAGCGGTGATGAGACTGCTCGAGGCTCCGAA 7200  
QY 2916 PheLeuPhePheLeuPheLeuHis 2923  
Db 7201 TTTCTCTTCTTAACTTCTGCT 7224

RESULT 6  
AX367102  
LOCUS AX367102 9321 bp DNA linear PAT 16-FEB-2002  
DEFINITION Sequence 21 from Patent W00198354.  
ACCESSION AX367102  
VERSION AX367102.1 GI:18855304  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Griffin,J.A., Kallick,D.A., Tribouley,C.M., Yue,H., Nguyen,D.B.,  
Tang,Y.T., Lal,P., Policky,J.L., Azimzai,Y., Lu,D.A., Grauli,R.,  
Yao,M.G., Burford,N., Hafalia,A.U., Baughn,M.R., Bandman,O.,

Patterson,C., Yang,J., Xu,Y., Warren,B.A., Ding,L. and  
Sanjanwala,M.S.  
TITLE  
Receptors  
JOURNAL Patent: WO 0198354-A 21 27-DEC-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..9321 /organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 6977010CB1"  
BASE COUNT 1780 a 3008 c 2776 g 1757 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 9321  
Score: 1958.00 Matches: 2886  
Percent Similarity: 96.81% Conservative: 0  
Best Local Similarity: 96.91% Mismatches: 17  
Query Match: 66.99% Indels: 92  
DB: Gaps: 0  
US-09-916-849a-3 (1-2923) x AX367102 (1-9321)  
QY 13 ProProProProLeuLeuLeuLeuLeuLeuLeuLeuLeuProProProLeuGlyAsp 32  
Db 262 CCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 321  
QY 33 GlnAlaGlyProCysAspSerLeuGlySerArgGlyYrYsArgGlyYrYsAlaCysAla 52  
Db 322 CAAGTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 381  
QY 53 PrometGlyTyrLeuCysProSerSerAlaSerAsnLeuTyrLeuYrYsThrSerArgCys 72  
Db 382 CCGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 441  
QY 73 ArgAspAlaGlyThrGlnLeuThrGlyYrYsLeuValProHisHisAspGlyLeuArgVal 92  
Db 442 AGGATGCGGGCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 501  
QY 93 TyrCysProGlnGlnGlnAlaHisIleProLeuProProAlaProGlyGlyCysProTyr 112  
Db 502 TGGGTCCAGATCCAGGCGCCATATTCCTTACACAGCTCTGAAAGGCTGCCCTGG 561  
QY 113 SerCysArgLeuLeuGlyIleGlyYrYsIleLeuSerProGlnGlyYrYsLeuThrLeuPro 132  
Db 562 AGCTGTGCTCTCTGGGCAATTGAGGCGCACTTTCCCAAGGGCAAGCTCAGACTGCC 621  
QY 133 GlnGlnHisProCysLeuYrYsAlaProArgLeuArgCysGlnSerCysYrYsLeuAlaGln 152  
Db 622 GAGGAGCAACCGTCTTAAAGGCTCCAGGCTCAGATGCCAGTCTCAGACTGGCAG 681  
QY 153 AlaProGlyLeuArgAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 172  
Db 682 GCGCCCGGGCTCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 741  
QY 173 ArgAsnValAsnThrAlaProGlnPheGlnProProSerTyrGlnAlaThrValProGly 192  
Db 742 AGGATGTAATACAGGCCCGCCCAAGTCCAGCCCGCCAGCTACAGGCAAGTCCGAG 801  
QY 193 AsnGlnProAlaGlyThrProValAlaSerLeuArgAlaIleAspProAspGlyGlyGly 212  
Db 802 AACCAAGCAGACAGGCAACCGTGTGCACTCCCTGAGGCGCATCCAGCCCGGAGAGGAG 861  
QY 213 AlaGlyYrYsLeuGlnGlyTyrThrMetAspAlaLeuPheAspSerArgSerArgGlnPhe 232  
Db 862 GCAAGTCACTGAGAGACCAATGATGCCCTCTTATAGCGCTCAACAGACTCTTC 921  
QY 233 SerLeuAspProValThrGlyAlaValThrThrAlaGlnGlyLeuAspArgGlyThrYrYs 252  
Db 922 TCCTGAGACCAAGTCACTGATGAGTAAACCAAGCGGAGAGGCTGATCGTGAAGACAG 981  
QY 253 SerThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeu 272

```
Db 982 AACACCCAGTCTTCAGGGTCACGGCGCAGAGACACGGCATGCCCCGAGAAAGTCCCTG 1041
Qy 273 AATThreuthrilleuvalThrAspThrAspAspHisAspProValPheGluGln 292
Db 1042 GCTACACTACCACTCTGGTTACTGACCAATGACCATGACCTCTGTTGAGCAGCAG 1101
Qy 293 GlnTyrGluSerLeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArg 312
Db 1102 GAGTACAAAGAGAGCTTCAGAGAGAACTGGAGTTGGCTATAGAGTGTCTCACTGCAAG 1161
Qy 313 AATThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySer 332
Db 1162 GCCACGAGATGGTATCCCTCCCAATGCAATATCTGTACCGCTGCTGAGGGGGTCT 1221
Qy 333 GlyGlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArg 352
Db 1222 GGGGGGAGCCCCCTCTAAAGCTTTGAGATGACCTCTGCGGGGTGATCCGAAACCGT 1281
Qy 353 GlyProValAspArgGluGluValGluSerTyrGluLeuThrValGluAspAspGln 372
Db 1282 GGGCCCTGGGATCGGGAAGAGTGGATCTTACCACTGACGGTACAGGCAATGACCAAG 1341
Qy 373 GlyArgAspProGlyProArgSerThrThrAlaIleValPheLeuSerValGluAspAsp 392
Db 1342 GGTCCGGAGCCCGGGGTCTCGAGTACACACGCGCTGTTTCTTCTGTGAGAGATGAC 1401
Qy 393 AsnAspAspAlaProGlnPheSerGluValArgTyrValValGluValArgIleAspVal 412
Db 1402 AATGAAATAGTCCCCCAAGTTTATGAGAAAGCTATAGTGTCCAGGTGAGGGAGAGTGG 1461
Qy 413 ThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAla 432
Db 1462 ACTCCAGGGGGCCCCAATACCTCCAGATCACAGCTCCGATGCAACAAAGGAGCAATGCC 1521
Qy 433 ValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGluPheTyrIleAspAlaGln 452
Db 1522 GTGGTCACTATAGCATCATGATGGCAATGCTGGGGAGCATGTTTATCTGATGAGCCAG 1581
Qy 453 ThrGlyAlaLeuAspValValSerProLeuAspTyrGluThrThrArgGluTyrThrLeu 472
Db 1582 ACTGGAGCTTGGATGTGTGAGCCCTCTTGACTATGAGACGACCAAGAGATCACCTTA 1641
Qy 473 ArgValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuValThr 492
Db 1642 CGGGTCGAGCAAGAGATGGTGGCCGTCGCCCACTCTTAATGTCTGTGGCTTGGTGA 1701
Qy 493 ValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAla 512
Db 1702 GTACAGGTCTGTGATATCAACGACAAATGCCCCCATCTTCGTACGACACCCCTTCCAGGCT 1761
Qy 513 ThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAla 532
Db 1762 ACTGTCTGGAGAGCGTCCCTTAAGCTACCTGGTCTTCAATGTCCAGGCTTACAGGCT 1821
Qy 533 AspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyIleAspPhePro 552
Db 1822 GATGTGTGTGACATGCGCCCTGGAATACCGCTTGTGGGGTGGAGATGACTTCCCC 1881
Qy 553 PheThrIleAsnAsnGlyThrGlyTyrPileSerValAlaAlaGluLeuAspArgGluGlu 572
Db 1882 TTACACATCAACAATGGCAGAGCTGGATCTGTGTGCTGTGAATCTGACCGGAGGAA 1941
Qy 573 ValAspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAla 592
Db 1942 GTTGATTTCTACAGCTTTGGGGTGAAGCTCGAGACCAATGGCACTCAAGCACTCACTGCC 2001
Qy 593 SerAlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGln 612
Db 2002 TCGGCGCAGTGTACGCTGCTGTGATGTCAACCAATCAATCAACTTAACTTAACTTAA 2061
Qy 613 ProGluTyrThrValArgLeuAsnGluAspAlaIleValGlyThrSerValValThrVal 632
Db 2062 CCAGAGTACACAGTGGCTCAATGAGATGCACTGTGGGCAACAGCGGTGTGACGGTG 2121
Qy 633 SerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThr 652
Db 2122 TCACTGTGAGACCGGATGCTCATAGTGTCACTACCTACAGATCAACAGTGGCAATACT 2181
Qy 653 ArgAsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuPro 672
Db 2182 CGAAACCCCTTCTCATCAACAGCCAAAGTGGTGGTGGCTGTATCTTCCCTTCCGCA 2241
Qy 673 LeuAspTyrGluLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArg 692
Db 2242 CTGAGCTCAAACTTGAACCGGCGATGATGTGGTGGCTGTATCCGCTCCGATGGCACTCGG 2301
Qy 693 GlnAspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisAspProValPhe 712
Db 2302 CAGAGACGGCACAAATGTGTGATGTCAACGAGCGCCACACCACTGCTCTTCTTT 2361
Qy 713 GlnSerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValVal 732
Db 2362 CAGAGCTCCCACTATACAGTGAATGTATAGAGACCGGCGGAGGACCAACGGGTGTG 2421
Qy 733 LeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGlu 752
Db 2422 CTGATCAAGCGCCACGAGTGAAGACACAGTGAATGCGCGATCACTTCAATGAG 2481
Qy 753 AspSerIleProGlnPheArgIleAspAlaAspThrGlyValAlaThrThrGlnAlaGlu 772
Db 2482 GACAGCATCCCACTTCGATTCGATCATGACACAGGGGGCTGTACCAACCGACTGAG 2541
Qy 773 LeuAspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIle 792
Db 2542 CTGAGCTACGAAAGACCAAGTGTCTTACACCTGGCCATTAAGTCTCGGGACATGCAATT 2601
Qy 793 ProGlnTyrSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAla 812
Db 2602 CCCCAAGAGTCCGACACCACTTACCTGAGATCTGTGGAGAGAGAGTGAATACAAATGCC 2661
Qy 813 ProGlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThr 832
Db 2662 CTTCAAGTCTTCGAGACCTCTTACCAAGGCAAGTGTATGAGAGATGTGCCACTTCACT 2721
Qy 833 SerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyr 852
Db 2722 ACGTCTCGCAATCTCAAGCACTGATCGTGAATTCGACCTTAATGCGAGGGCTCTTAC 2781
Qy 853 ThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIle 872
Db 2782 ACCTTCGAAGAGCGACATGAGACGGTGACTTATGTGTGAGTCCAGTCAAGGCATC 2841
Qy 873 ValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyr 892
Db 2842 GTGGCAACCTTACGAGAGCTGGATCGAGAGACGTGGCCCAATATGTCTTGGGGCATAT 2901
Qy 893 AlaValAspGlyGlyMetProProAlaArgThrProMetGluValThrValThrValLeu 912
Db 2902 GCAGTGACAAAGGGATGCCCCCAGCCGCAACCTATGGAAGTGCACATGCTGTGTG 2961
Qy 913 AspValAsnAspAsnProProValPheGluGluAspGluAspValPheValGluGlu 932
Db 2962 GATGTGAATGACATCCCTCTGTGAGAGATGAGATGAGTGTGTGTGTGAGAGAG 3021
Qy 933 AsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThr 952
Db 3022 AACAGCCCATTTGGGTAGCGTGGCGGGTCAAGCCACTGACCCCGATGAGGCAACC 3081
Qy 953 AsnAlaGlnIleMetTyrGlnIleValGluGluAsnIleProGluValPheGlnLeuAsp 972
Db 3082 AATGCCCAATTAATGACAAATGTGAGGAGCAACATCCCTGAAGGTCTTCCAGCTGAGAC 3141
Qy 973 IlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyr 992
Db 3142 ATCTTCTCCGGGAGCTGACACCCCTGTGATGACTTATGACTACGAGGACCGGCTGATGAC 3201
```

QY 993 ValLeuVal11IeGlnAlaThrSerAlaProLeuValSerArgAlaThrValHISValArg 1012  
DB 3202 GTCCGTGATCATCGAGCCACGTCAGCTCTCTGTGTAGCCGGGCTACAGTCCACGTCGCC 3261  
QY 1013 LeuLeuAspArgAsnAspAsnProProValLeuGluYAsnPhenGlnu1LeuPheAsnAn 1032  
DB 3262 CTCTCTGACCGCAATGACAACCAACGAGCTGGGCAACTTGAAGATCTTTTCAACAAC 3321  
QY 1033 TyrVal1ThrAsnArgSerSerPheProGluYAla1IeGluYArgValProAlaHIS 1052  
DB 3322 TATGTACCAATCGCTCAAGAGCTTCCCTGGGGGGGCCATTGGCGGATTACTGCGCAT 3381  
QY 1053 AspProAsp1IeSerAspSerLeuThrYrSerPheGluArgGluYAsnGluLeuSerLeu 1072  
DB 3382 GACCTGTAAATCTCAAGATGTCTGACTTACAGCTTGAACGGGGAATACCTCAAGCTCG 3441  
QY 1073 ValLeuLeuAsnAlaSerThrGluYLeuYLeuSerArgAlaLeuAspAsnAnArg 1092  
DB 3442 GTCTGTCTCAATGCTTCCACGGGTAGCTGAAGCTGAAGCCGCACTGGACAACAACCGG 3501  
QY 1093 ProLeuGluAla1IeMetSerValLeuValSerAspGluYAlaHISerVal1ThrAlaGln 1112  
DB 3502 CCTGTGAGGGCCATATAGCGGTGTGTGAGACGGGTACAGAGCTGAACGCCACG 3561  
QY 1113 CysAlaLeuArgVal1Thr1Ie1ThrAspGluMetLeuThrHISer1IeThrLeuArg 1132  
DB 3562 TGGCGGCTGGGTGTGACATCAACCGATGAGATGCTCAACCAACGATCAACGCTGCGC 3621  
QY 1133 LeuGluAspMetSerProGluArgPheLeuSerProLeuLeuGluYLeuPhe1IeGlnAla 1152  
DB 3622 CTGAGAGACATGTACCCGAGGGCTTCGTGACCACTGAGGCTCTTCAATCCAGGGG 3681  
QY 1153 ValAlaAla1ThrLeuAla1ThrProAspAlaValAla1PheAsnValGlnArgAsp 1172  
DB 3682 GTGGCGCGACGCTGGCCACCGCACCGAGCCAGTGTGTCTTCAACGTACAGCGGAGC 3741  
QY 1173 ThrAspAlaProGluYAlaHIS1IeLeuAsnValSerLeuSerValGluYLeuProGluY 1192  
DB 3742 ACCGAGCGCCCGGGGGGCGACATCTTCAACGTGAGCTGTGGTGGGCGACGCGCAGGG 3801  
QY 1193 ProGluYAlaYProProPheLeuProSerGluAspLeuGlnuArgLeuYrLeuAsn 1212  
DB 3802 CCGCGGGGGGGCGCGCCCTTCTGCTCTGAGACCTGACAGAGGCGCTAATACCTCAAC 3861  
QY 1213 ArgSerLeuLeuThrAla1IeSerAlaGlnArgValLeuProPheAspAsn1IeCys 1232  
DB 3862 CGCAACCTGTGACGGGCATCTCGGACAGCGCGGTGTGCTTGAAGACAACATCTGC 3921  
QY 1233 LeuArgGluProCYGluAsnTYrMetArgCYsValSerValLeuArgPheAspSer 1252  
DB 3922 CTGCGGAGGCGCTGCGAGAACTACATGCGCTCGTCCGTCCGTCCGCTTCACTCTCC 3981  
QY 1253 AlaProPhe1IeAlaSerSerSerValLeuPheArgPro1IeHISProValGluYLeu 1272  
DB 3982 GCGCCCTTCATGCTCTCTCTCGGCTCTTCCGCGCATCAACCCGCTGGAAGGGCTG 4041  
QY 1273 ArgCYsArgCYsProProGluYrPheThrGluYAspTYrCYsGluThrGluValAspLeuCYs 1292  
DB 4042 CGCTCCGCTGCGCGCGCGGTTCACGGGTGACTCTGTGAGACCGAAGGTGAACCTCTGC 4101  
QY 1293 TyrSerArgProCYGluYProHISgluYArgCYsArgSerArgGluYgluYrThrCYs 1312  
DB 4102 TACTGCGCGCGCTGTGGCCCGCACGGGCGCTGCCACGCGCGAGGGGGGCTACACCTGC 4161  
QY 1313 LeuCYsArgAspGluYrThrGluYHISCYsGluValSerAlaArgSerGluYArgCYs 1332  
DB 4162 CTCTCTGTGATGCTACACGGGTGAAGCATGTGAAGGTAGTGTGCTGCGCTCAAGCGGTGC 4221  
QY 1333 ThrProGluYAlaCYsAlaAsnGluYgluYrThrCYsValAsnLeuLeuValGluYrPheYs 1352  
DB 4222 ACCCGGGGTGTCTGCAAGAAATGGGGGCACTGTGTCAACTGTCTGTGGGCGGTTTCAAG 4281

QY 1353 CysAspCYsProSerGluYAspPheGluYsProTYrCYsGlnVal1Thr1ThrArgSerPhe 1372  
DB 4282 TGGATGCGCATCTGAGACTTGCAGAAAGCCCTTACTGCGAGGTGACCAACGGGAGCTTC 4341  
QY 1373 ProAlaHISerPhe1IeThrPheArgGluYLeuArgGlnuArgPheHISerThrLeuAla 1392  
DB 4342 CCGGCGCACTCTTCATACCTTTTGGCGGCGCTGCGCGCGCTTTCACCTTCACTTCCGCGC 4401  
QY 1393 LeuSerPheAla1ThrGluYAspGluYLeuLeuLeuYrAsnGluYArgPheAsnGlu 1412  
DB 4402 CTCTGCTTGGCCACAAGAGAGCGCACGGGTGTGCTGTGTACAAATGGCGCTTCAATGAG 4461  
QY 1413 LysHISAspPheValAlaLeuGluVal1IeGlnGluValGlnLeuThrPheSerAla 1432  
DB 4462 AAGCATGACTTGTGTGGCCCTCGAGGTGATCCAGAGAGCTTCAACTCACTTCTGTGA 4521  
QY 1433 GluYIleSerThrThrThrValSerProPheValProGluYAlaSerAspGluYLeu1Trp 1452  
DB 4522 GGGAGTCAACCAACGAGGTGTCCCATTTGTCGCCGAGAGATCAATGGCCCAAGTGG 4581  
QY 1453 HISThrValGlnLeuYsTYrTYrAsnLYsProLeuLeuGluYGlnThrGluYLeuProGln 1472  
DB 4582 CATACGGTGCAGCTGAATATACAAATAAGCACTTTGGGTCAAGACAGGCTCCCAAG 4641  
QY 1473 GluYProSerGluGlnYsValAlaValAlaVal1ThrValAspGluYCYsAspThrGluYAla 1492  
DB 4642 GGCCCATCAGAGCAGAAAGGTGGTGTGTGACCGGTGATGCTGTGACACAGAGAGTGGCC 4701  
QY 1493 LeuArgPheGluYsValLeuGluYAsnTYrSerCysAlaAlaGlnGluYrThrGlnGluY 1512  
DB 4702 TTGGCTTGGATCTGTCTGGGCACTACTCTGTGTGTCACAGGGGCAACCAAGGTGGC 4761  
QY 1513 SerLYsLYsSerLeuAspLeuThrGluYProLeuLeuLeuGluYAlaProAspLeuPro 1532  
DB 4762 AGCAAGAGTCTGTGANTCTGACGGGGCCCTGTGTACTAGCGGGGTCTGACCTGCGC 4821  
QY 1533 GluSerPheProValArgMetArgGlnPheValGluYCYsMetArgAsnLeuGlnValAsp 1552  
DB 4822 GAGAGCTTCCAGTCCGAAATCGGACAGTTCGTGGGCTGCAAGCGGAACTGACAGGTGAC 4881  
QY 1553 SerArgHIS1IeAspMetAlaAspPhe1IeAlaAsnAnGluYrThrValProGluYCYsPro 1572  
DB 4882 AGCGGACATAGACATGAGCTGACTTCAATTCACAAATGGCAACGTCGCTGGCGCT 4941  
QY 1573 AlaLYsLYsAsnValCYsAspSerAsnThrCYsHISAsnGluYgluYrThrCYsValAsnGln 1592  
DB 4942 GCAAGAGAGAGCTGTGAGACGACACACTTGCACAAATGGGGGCACTTGTGTGAACAG 5001  
QY 1593 TrpAspAlaPheSerCYsGluCYsProLeuGluYrPheGluYgluYrSerCYsAlaGlnGlu 1612  
DB 5002 TGGACCGGTTCACTGCAAGTGTCCCTGGGCTTGGGGGCAAGAGCTGCGCCACAGAA 5061  
QY 1613 MetAlaAsnProGlnHISpHeLeuGluYSerSerLeuValAlaTrpHISgluYLeuSerLeu 1632  
DB 5062 ATGGCAATCCACAGCACTTCTGGGACAGAGCTGTGGGTGACATGAGGCTCTCGGTG 5121  
QY 1633 Pro1IeSerGlnProTYrTYrLeuSerLeuMetPheArgThrArgGlnAlaAspGluYAla 1652  
DB 5122 CCAATCTCCCAACCTGTGACTCAAGCTCATGTTCGCAAGCGCCAGCGGAGGAGTGT 5181  
QY 1653 LeuLeuGlnAla1IeThrArgGluYArgSerThr1IeThrLeuGlnLeuArgGluYHIS 1672  
DB 5182 CTGCTGACGGCCATACACAGGGGGCGAGCACATCAACCTTACAGCTTACAGAGGGCCAC 5241  
QY 1673 ValMetLeuSerValGluYrThrGluYLeuGlnAlaSerSerLeuArgLeuGluYProGluY 1692  
DB 5242 GTGATGCTGAAGCTGAGAGGGCAAGGGCTTCAAGGCTCTCTCTCTGAGGCAAGGC 5301  
QY 1693 ArgAlaAsnAspGluYAspTrpHISAlaGlnLeuAlaLeuGluYAlaSerGluYgluYrPro 1712  
DB 5302 CGGCGCAATAGCGGTGACTGGCAACATGCAACGTGGCACTGGAGCGAGGGGGGCT 5361  
QY 1713 GlyHISAla1IeLeuSerPheAspTYrGluYLeuGlnAlaArgAlaGluYAsnLeuGluYPro 1732

Db 5362 GGGCAAGCCATCTCTGCTTCGATTATGGGACAGAGAGAGGGCAACCTGGGCCCC 5421  
 Qy ArgLeuHieglYleuHieSerAamIleThrValGlglyIleProglYProIleaglY 1752  
 Db 5422 GGGCTGCAAGGTGTGACCTGAGCAACATACAGTGGCGGAATACCTGGGCCAGCGCG 5481  
 Qy 1753 GilyValAlaarglyPhearglyCybleugInglyValaIargValSerAapThrProglu 1772  
 Db 5482 GGTGTGGCCCGTGGCTTTGGGGCTGTTTGCAGGGTGTGGGGTGGAGCCATAGCCAGAG 5541  
 Qy 1773 GilyValAlaSerIleuAapProSerHieglYgluSerIleAamValIgluInglyCySer 1792  
 Db 5542 GGGGTTAACAGCTGATCCACAGCATGGGAGAGCATCAACGTGAGCAAGCTGTACG 5601  
 Qy 1793 LeuProAapProCyAapSerAapProCyAaIleAamSerTyCySerAaAapThr 1812  
 Db 5602 CTGCTGACCTTGTGATCAACCCGTGCTGTGCTACAGCTATGTGACAGCAAGACTGG 5661  
 Qy 1813 AapSerTySerCySerCyAapProglYTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 1832  
 Db 5662 GACAGCTATTCTGACAGCTGTATCCAGGTATCTATGTGACAACTGTATATGTGTGT 5721  
 Qy 1833 AapLeuAapProCysgluHiegluSerValCyethrArglyAapProSerAlaProHieglY 1852  
 Db 5722 GACCTGAACCCGTGTGAGCAACAGCTGTGTGTACCCGCAAGCCAGTGGCCCCCATGGC 5781  
 Qy 1853 Ty 1872  
 Db 5782 TATACCTGCGAGGTGTCCCAAAATACCTTGGGCCATACGTGTAGACAGAAATTAACAG 5841  
 Qy 1873 ProCyAapProglYTy 1892  
 Db 5842 CTTGTGCCCCGTGGCTGTGTGGGACATCCCAATGTGGCCATGACACTGTATGTGTCAAG 5901  
 Qy 1893 Ty 1912  
 Db 5902 AAAGGCTTTGACCAACATGCAACAGAGGCGGAGTGCACATGCAAGGAGAAACCAAC 5961  
 Qy 1913 Ty 1932  
 Db 5962 TACCGGCCCCCAGGACAGCCCACTGCTCTTGTGTGATCTGTATCCCAAGGCTCTTGG 6021  
 Qy 1933 SerArgValCyAapProgluAapglYgluCyAaProCyAaIleAapProglYValIleaglYArg 1952  
 Db 6022 TCCAGAGTCTGTGACCTGAGATGGCAGGTGTCCATGCAAGCCAGGTGTATCGGGCGT 6081  
 Qy 1953 GluCyAapArgCyAapAapProPheAlaIleValIleThrArgSerIleCySerIle 1970  
 Db 6082 CAGTGTGACCGCTGTACAACTTTGTGAGGTGACCAACAAATGCTGTGAAAGGGCCC 6141  
 Qy 1971 --ValAamTyraPser--CyAaProAlaIle--Glu--AlaGlyIleThrTyProArg 1988  
 Db 6142 TTTGTTTG-CT-AG-TT-ACGTGTCCCG-GCCCA-TGAAGTGTGGCC--TC-CA-GC-AG 6190  
 Qy 1988 gthr-----Arg--PheglY-----LeuProAla-AlaAlaProCyAaPr 2000  
 Db 6191 AACCTGTACAGCTGTACAGGGGCTTCCT-GTGTGTCCCTGA-GGCCGG-CCCTT-TT 6246  
 Qy 2000 olySglYserPhe-----GlyThraIaValaArgHieCyAapgluHieAarglyTyTyTyTy 2018  
 Db 6247 -GG-CT-TCTTTCCTCCCAAGGAGCTGTGTGGCTGTATATAGCAAGGGGGTGTGCTC 6303  
 Qy 2019 ProProAamLeuPheAamCyThrSerIleThrPheSerIleuTySglYPheAlaIle 2038  
 Db 6304 CCCCACAACTCTTCAACTGCACTGCATCACTTCTCAAGAACTGAAGGGCTTCGTGAG 6363  
 Qy 2039 ArgLeuGlnArgAangIuSerIleYleuAapSerGlyArgSerGlnIleuAlaIleuLeu 2058  
 Db 6364 CCGCTACAGGAGATGTGTAGGCTAGACTCAGGGGCTCCCAAGAGCTAGCCGTGCTC 6423  
 Qy 2059 LeuArgAamAlaThrGlnIleThrAlaGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 2078

Db 6424 CTGCGCAACGCCACGCAACAGCTGTACTTGGCAGGAGCTCAAGGTGGCTTAC 6483  
 Qy 2079 GlnLeuAlaThrArgLeuLeuAlaHiegluSerThrGlnaArgGlyPheglYleuSerAla 2098  
 Db 6484 CAGTGGCCACCGGGCTGTGGCCCAAGAGCAACCCAGGGGGCTTTGGGCTGTGTGCC 6543  
 Qy 2099 ThrGlnAapValHiePheThrGluAamLeuAargValIglYserAlaIleuLeuAapThr 2118  
 Db 6544 ACACAGAGGTGCACTTCACTGAGAAATCTGTGGCGGGTGGGAGCGGCTCTGTGACACA 6603  
 Qy 2119 AlaAamTyraArgHieThrGluLeuIleGlnIleThrGlnuIleGlyTyThrAlaThrPheLeu 2138  
 Db 6604 GCCAACAGCGGCACTGTGAGCTGATCCAGACAGAGAGGGTGGACCGCTGTGTGCTC 6663  
 Qy 2139 GlnHieTy 2158  
 Db 6664 CAGCACTTAGGCTTACGCCAGTGCCTTGGGCCCAAGCATGGGCAACCTTACCTTACG 6723  
 Qy 2159 ProPheThrIleValIleThrProAamIleValIleSerValaIargLeuAapIleYasn 2178  
 Db 6724 CCTTACACATGTGACGCCCAACATGTGTCACTCTCGTAGTGGCTTGGAACAAAGGAAC 6783  
 Qy 2179 PheAlaIleValaIleuLeuProArgTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 2198  
 Db 6784 TTTGTGTGGGCGCAAGCTGCCCGCTTACAGAGGCTGTGTGGAGAGACCCCGGACCTT 6843  
 Qy 2199 GluThrThrValIleLeuProgluSerValPheArgGluThrProProValaIleAapPro 2218  
 Db 6844 GAGACAAAGTATCTGTGCTGTAGTCTGTCTTCAAGAGAGGCCCCCGTGTGACGGCCC 6903  
 Qy 2219 AlaIleProglYgluAlaIleGlnIleProgluIleuAlaIargArgIleAapPro 2238  
 Db 6904 GCGAGCCCGGAGAGGCGCCAGGAGCGAGAGGCTGGCAGCGGCGACAGCGGACCGC 6963  
 Qy 2239 GluLeuSerGlnIleGluAlaValaIleSerValIleIleTyTyTyTyTyTyTyTyTyTy 2258  
 Db 6964 GAGCTGAGCCAGAGGTGAGGCTGTGGCCAGCTCATCATCTACCTGAGCCGAGGCTA 7023  
 Qy 2259 LeuProHieAamTyraAapProAapIleAapIleAapIleAapIleAapIleAapIle 2278  
 Db 7024 CTGCTCATTAATTAAGACCTGTGACAGGCGAGCTTGAAGAGTCCCAAGCCGCAATCATC 7083  
 Qy 2279 AamThrProValIleSerIleSerValHieAapAapgluIleuLeuProArgAlaIleu 2298  
 Db 7084 AACACACCCGTGTGAGATCAGCTGCATGATGATGAGGCTTCTGCCCCGGGCGCTG 7143  
 Qy 2299 AapIleProValIleThrValGlnPheArgLeuIleuGluThrGlnuIleAargThrIle 2318  
 Db 7144 GACAAACCCGTGACGCTGAGTCCGCTGTGAGACAGAGAGCGGACCAAGCCATC 7203  
 Qy 2319 CyValaPheThrAamHieSerIleuValSerGlyTyThrGlyTyTyTyTyTyTyTyTyTyTy 2338  
 Db 7204 TGTGTCTTGTGAACATTCATCACTGTGTCAAGTGCACAGGTGGTGTGGCGGAGAGG 7263  
 Qy 2339 CySgluValaIlePheArgAangIuSerHieValSerCySgluCyAamHieThrSer 2358  
 Db 7264 TGTGAAGTGTCTTCCGAAATGAGACAGCGCAAGCTGCAGTGCACACATGACGAGC 7323  
 Qy 2359 PheAlaValleuMerAapValSerArgArgIle 2369  
 Db 7324 TTTGCTGTGCTATGAGAGTGTCTGTGGCGGAGGTGGGCCCAAGGGGAGCTGCAGAG 7383  
 Qy 2370 -----AangIyGluIleuProleuTyThrLeuThrTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 2387  
 Db 7384 CCGTGAATGGGAGATCTGTGCACTGAAGACACTGATAGTGTGTGAGTGTCAAC 7443  
 Qy 2388 LeuAlaAlaIleuLeuLeuThrPhePheLeuThrIleuLeuAargIleIleuAargSerIle 2407  
 Db 7444 TTTGGTGGCTTGTGTCACTCTTCTTCTCACTCTTGTGATCTGTGGCTGTGCTCAAC 7503  
 Qy 2408 GlnHieglYIleArgAamLeuThrAlaAlaIleuGlyIleuAlaIleuValPheLeu 2427  
 Db 7504 CAACAGGATCCAGATCACTACAGCTGCGCTGGGCTGTGCTCACTGTCTTCTC 7563



QY	2428	LeuGLy11LeaNGlnAlaSPLeuProPheAlaCySrrtVal11LeuAl1LeuLeuH1s	2444
Dh	7564	CTGGAAATCAACACAGGCTGACCTCCCTTTGGCTGCACAGTCAATTGCCATCCTGCTGCAC	7622
QY	2448	PheLeuTYrLeuCySrrtPheSerTYrAlaLeuLeuGlnAlaLeuH1sLeuTYrArgAla	2466
Dh	7624	TTCCGTACCTCTGCACCTTTTCCGGGCTCTGCTGGAGGCTTGGACCTGTACCGGGCA	7683
QY	2468	LeuThrgLlValArgASPValAsnThrgLYProMeCArGpHeTYrTYMeLeuG1YTrp	2487
Dh	7684	CTCACTAGAGTGGCGGAGTGTACACACCGGCCCAATGCCCTTCTACTACATGCTGGGCTGG	7743
QY	2488	GLYValProAlaPhe11eThrgLYLeuAlaValGlyLeuASPProGlnLYTrG1YAsn	2507
Dh	7744	GGCGGCTGCTTCATCATCACAGGGCTACCGGGGCTGGACCCCGAGGGCTACGGGAAC	7803
QY	2508	ProASPheCySrrtPLeuSer11eYrAspThrLeu11eTrpSerPheAlaG1YProVal	2527
Dh	7804	CTTACCTTCTGCTGGCTCTCCATCTATACACGCTCATGTGAATTGTCTGGCCCGGTG	7863
QY	2528	AlaPheAlaValSerMetSerValPheLeuTYr1LeuAla1AlaArgAlaSerCyAla	2547
Dh	7864	GCTTTGGCGCTCGATAGTGTCTCTCTGTACACCTCGGGGCGCCGGCTCTGTGTCT	7922
QY	2548	AlaGlnArgGlnGlyPheGlnLYsGlyProValSerGlyLeuGlnProSerPheAla	2567
Dh	7924	GCCACAGCGGACGGGCTTGGAGAAGAAAGTCTGTGCTGGGCTGTACGCCCTCTTGCC	7983
QY	2568	ValLeuLeuLeuLeuSerAlaThrThrPLeuLeuAlaLeuLeuSerValAsnSerThpThr	2587
Dh	7984	GTCTCTCCGCTGCTGAGGCGCACGCGGTGGGACGTCTGTGTACACAGGACCAACC	8043
QY	2588	LeuLeuPheH1sTYrLeuPheAlaThrCyAsnCyS11eGlnGlyProPhe11ePheLeu	2607
Dh	8044	CTCTCTTCCACTACCTCTTTGGTCTACCTGTACATTCACACAGGCGCCCTTCACTCTCTC	8103
QY	2608	SeTYrValValLeuSerLYsGlnValArgLYsAlaLeuLYsLeuAlaCySerArgLYs	2627
Dh	8104	TCCATATGGTGGCTTACGACAGAGGTCCGGAAGACCTCAAGCTTGCTGCAGCGGCAAG	8163
QY	2628	ProSerProASPProAlaLeuThrThrLYsSerThrLeuThrSerSerTYrAsnCyPro	2647
Dh	8164	CCCAAGCCCTGACCTGCTGTACACCAAGTCCACCCCTGACCTGCTGTACACCTGCC	8223
QY	2648	SeTYrTYrAlaAspG1YArgLeuTYrGlnProTYrG1YAspSerAlaG1YSerLeuH1s	2667
Dh	8224	AGCCCTTACGAGATGGGCGGCTGTACACGACCTTACGAGACTGGCGGCTCTGTGCAC	8283
QY	2668	SeThrSerArgSerG1YLYsSerGlnProSerTYr11ProPheLeuLeuArgG1uGln	2687
Dh	8284	AGCACACAGTGCCTCGGGCAAGGTCCAGCCAGCTACATCCCTTGTGTACGGGAGGAG	8343
QY	2688	SerAlaLeuAsnProG1YGlnGlyProProG1YLeuG1YAspProG1YSerLeuPheLeu	2707
Dh	8344	TCCCACTGAACCTCGGCGCAAGGGCCCCCTGGCTGGGGAGTCCAGGACGCTGTCTCTG	8403
QY	2708	GLYGLYGLNAspGlnGlnH1sASPProAspThrAspSerAspSerAspLeuSerLeuGln	2727
Dh	8404	GAACTGTACAGCCACAGACATGATCTTCAACAGGACTCCACAGTACCTGTCTTGTAA	8463
QY	2728	ASPAspGlnSerG1YSerTYrAlaSerThrH1sSerSerAspSerGlnG1uGlnGln	2747
Dh	8464	GACGACCAAGATGGGTCTTATGTCCCTTACCCACTCATCAGCATGTAGAGAGAGAGAG	8523
QY	2748	GLYGLYGLYGLYGLYAla1AlaPheProG1YGLYGLYGLY1YTrpAspSerLeuLeuG1YPro	2767
Dh	8524	GAGAGGAGAGAGGAGGCGCTTCTCTGTGAAGCAAGGCTGGATAGCTCTCTGGGCGCT	8583
QY	2768	GLYAlaGlnAspGLeuProLeuH1sSerThpProLYsAspG1YGLYProG1YProG1YLYs	2787
Dh	8584	GGAGCGAGAGACTGCCCTCTGCACATCTCCACAGATGGAGGCGCCACGGGCTGTGGAG	8643

Oy	2788	AlaProIrrPProGIYAAPPhaeGIYThThAlaYseGIuSeSerGIYAseGIYAlAPro	2807
Db	8644	CCCCCTGGCCAGAGACTTTGGGACCAACAGCAAAAGAGTAAGTGGCACGGGGCCCTT	8703
Oy	2808	GIuGIuAryLeuAryGIuAseNGIYAseAlaLeuSeAryGIuGIuGIYSeLeuGIYProLeu	2827
Db	8704	GAGAGAGGGGCTGGCGGAGAAATGGAGATGCCCTGCTCGAAGAGGGGTCCCTAAGGCCCTT	8753
Oy	2828	ProGIYSeSerAlaGIuInProHAlaYseGIYTLelenuYelYseCySeLeuProThTlle	2847
Db	8764	CCAGGCTCTTGCCACGCTCACAAAGGCAATCTTAAAGAAAGATGTCTGGCCACCAATC	8823
Oy	2848	SeGIuIYSeSerSeuLeuAryLeuProLeuGIuGIuGIYCyThrGIYSeSerAryGIY	2867
Db	8824	AGCGAAGAGACAGACCTCTCGGGCTCCCTCGAGCAATGACAGAGGTCTTCCGGGGGC	8883
Oy	2868	SeSerAlaSeerGIuGIYSeAryGIYGIYProProProAryProProAryGIuInSer	2887
Db	8884	TCTCTCCCTAGTAGAGGCAACCCGGGGGGCCCCCTCCCGCCACCGCCCCGGCAAGC	8943
Oy	2888	LeuGIuGIuGIuInAseNGIYAlaMetProIleAlaMetSeuLeuYAlaGIYThVal	2907
Db	8944	CTCCAGAGAGCACTGAACGGGGGTCAATCCCATCGCATGACATCAAGCAGCACGGTG	9003
Oy	2908	AseGIuAseSeerGIYSeGIYSeGIY 2915	
Db	9004	GATAGGAGCTCGTCAAGCTCCGAA 9027	
RESULT 7			
AX646739			
LOCUS	AX646739	24370 bp	DNA linear PAT 04-MAR-2003
DEFINITION	Sequence 931 from Patent EP1270724.		
ACCESSION	AX646739		
VERSION	AX646739.1	GI:28799122	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			
AUTHORS	Swia, M., Asai, K., Akiyama, Y. and Aburatani, H.		
TITLE	Guanosine triphosphate-binding protein coupled receptors		
JOURNAL	Patent: EP 1270724-A 931 02-JAN-2003;		
	National Institute of Advanced Industrial Science and Technology		
	(NP) ; Center for Advanced Science and Technology Incubation, Ltd.		
	(NP)		
FEATURES			
source	Location/Qualifiers		
	1..24370		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
CDS	join(201..3510,8553..9200,11163..11385,11634..11749,		
	11929..12017,12408..12565,12927..13090,13263..13388,		
	13732..13898,14199..14486,14575..14749,14969..15130,		
	15312..15432,15856..16039,16226..16327,17669..17763,		
	17972..18181,18702..18885,19001..19129,19231..19411,		
	19545..19712,19815..19941,20054..20254,20547..20721,		
	21048..21160,21338..21417,21507..21633,21721..21843,		
	22399..22526,22761..22849,22954..23148,23287..23457,		
	23557..23791,24143..24170)		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD69495.1"		
	/db_xref="GI:28799123"		
	/translation="MRSPATGVLPFPPEPLILILILPPLLDQGVGCRSLGSRG		
	RGGSGCAPMGWLCPSSANLMLYTRCRDAGTELGHVPHNDGLRWCPESBAH1P		
	LPAEESCPWSSCRLLIGTISPOGKLTLYPEHPCIKAPLRCOSCLDAPGLRAGE		
	RSPEISGGRKKNVNTAIPQFOSYAATVPEHNPAGTPVASRALDIPDEGAGREY		
	TMALPDSEHNOFSLDPTVGAATTAEBELREIKRHVPTVADODHMGKALATLT		
	LIATPDNDPVEOCEYKESAEINLEUGTEVLYATATGDAAPNNILIRILEGSG		
	SPSEVEIDRSGVITIRGPDVDEBEYESTYLVEASDQGDPEPRSTAAVFLSVSD		
	NDNAPFSESRVYVQVREIDTPGAPVLAIRVLAASRDGKSNVHYSTINSAGQFYLD		
	AQCGALDVSPLEDTETTKETKVTAVRQDGGRPPLSNVSGLVATVOVIDNNAQIPVST		

PFOATVLESVPLGTVLTHVQALDADAGDNARLETRLAGVHDPEPTTINNGTWNISVA  
 ELDEBEVDFYSFGEARHDGTGTPALTPALTSASVTVLVDNNPFPOTPEVTLEREDYA  
 VGTSVTVASVVDADASVITTOITSGNTRRPSITSOSGGGALSLPYDLKEREDYA  
 LATVADGTRODFOAOIVVANTDANTHRPPOSHTYVNSHTYVNTDPACTYVLSANDED  
 TGNMATITTYMESTIPQFRIDADTCAVTTQALDYEDQVSYTLATITARDNGIPKSDT  
 TYLILVNDVNDNAQFPLDSDYQGSVEDEPPTSVLQISADRDGSLNGRVEYTOG  
 GDGDDDFIVESTSGIVRTLRDRNVAQVYARAVAKMGPAPTEMEVTVLVD  
 NNDPVEDEDPVPEBNSPIGLAVARVATDDEGTQAQIYQIVGNEIPEVQLD  
 IFSGELTALVDLDEDRPEXVLIQNTSAPVSRATVATVRLDRNPNPLVNGEFLP  
 NNVTNRSSPFGGAIQVPAHPDIDSLITVSEFERNGLSDILVNAQSLKASAL  
 DNNRPIEALNSVVSQSHSVTAQCALRTIITDENLTHSITLRLEDMSPERFLSL  
 GLTQAVATLATPEOHVAVFNQDNDIAFGHILNVSLSVQPPGPGRGAPPEYFED  
 LQERLVNLSLTLAISORVLPDNDICLRBPCENYMRCVSYRPFPSAPFLSSVTL  
 FRPIHVGQGLRCRCPGFTDCEYEDVLCYRCPGPHGCRREGGYLCPASHEFT  
 EHCESVARSGRCTPGYKNGGTQCNVLLVGGFKDCPSGDDEKRYCOVTTNRSPASHFI  
 TTRGRLORPHFTLASPTATKERDGLLNNRERKHPVALATVIOBNOLOTPBAGST  
 TTVSPVYPCGVSPGQHTTQOLKTKYKPLAQTCGLPGPSEOKAVATTVDCGTQALR  
 PGLVANTSCAQGTQGSKSLDTGLPLLGVDLPSPSPVNRMOQVPGKMENTLVD  
 SRHIDMADFLANNQVPGCPAKNVDSDNTCHNGTQVNMQDAPSCBPGFSGKSCA  
 OEMANPOHFLGSSLVAMHGLSLPISQPMYSLMFRROADGVLLQAITRGRSTTLQ  
 REGHVMLSYEGTLOASSLRLEBRANDGMHHAOLALAGSGGHALISPDYGOORA  
 BGNLGRHLGHLISNITVGGIPGAGVARGPFGICQYRVSPTPEGVNSLDSHRS  
 INVOCGSLPDCDSNCPANYSICSDMDSYSCSCPRTYGDNCTVNCILNPEHOSV  
 CTRKESAPHGTCPCPNVLDGTCERTIDQPCRGMMGHPTGCPNCNDVSKGDPDCN  
 KTRSGECKENHYRPPSPCLCDYPTGSLRVCDPBDQCPCKRQVYGRCDERCD  
 NPEAVTYTNGCEVYVDSCPRAIEAGIWPETRRGLPALAACPGYSFGTAVRCDERCD  
 WLPENLNTCTSYFSELKGPABERLORNESGLDSGRSOOLALRNATOHATGYGSDV  
 KVALATATRLAHESTORGGLSATODVHTTELTVGSLALDTRKRMWELILOQEG  
 GRATVLOHYEAVASALAKONRHTYSPPTITVTNITVYSYVRLDKGNPAKAPRYTAL  
 RGSQPDLETTVILPESVRETPPVVRPAGPSEAOPREBARORRHELSQSEAVAS  
 VILYRTLAGLPHNDYVNDKSLNVPKRPINTVVASIYVHDEBLPLRDLRPNVTVQ  
 RLLETERTPCICVFNWHSILVSGTGMARSCEVFRSESHSCCNMTSPASVAMD  
 VSRERENGLPLKTLTYVALAGTVALLTFPFLTLILRSOGLIRMLTALALGIA  
 OLVTFLGINOALDLPACTVAILLHFLVICTPFWALLBALHLYALITERDVTGMBR  
 FYMNGMGPARTGLANGLDPRGYNPPTCMISTYDTILNSAGVYARAVSKSVLY  
 ILAARSCAQRQGEKKEPVSGLQSPFAVLHLSATWLSLVSVDLTFLELAT  
 CMCIDGPFPLSTVYLSKEVRKALKACSKRKPDPALTKTSLTSYNSPCFSTYADGR  
 LYOPYDGSAGSLSTDSRSGKSPSYIPLILREBSALNPGCPGLGDPSLFLQEGDO  
 OHDPDSDSDSLIEDDQSGYASTHSDBSEBEEBEEBEEBAAAPGQMDSLIGPAP  
 RLPLSTPKDGGPGKAPWPDGRTTGAATKSSGNGAPEERLRNGALRREGSLGLP  
 GSSAOPHKGLIKKKCLPTISEKSLRLRPLEOCTGSSRSASBESRGCPRRPPRRPO  
 SLOBQUNGMPIMSTIKACTVDEDSGSEFLFPNPLH"

BASE COUNT 4543 a 7241 c 7458 g 5128 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 24370  
 Score: 1103.00 Matches: 1103  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.74% Indels: 0  
 DB: 6 Gaps: 0

US-09-916-849a-3 (1-2923) x AK646739 (1-24370)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 DB 201 ATGGGAGCCGGGCAACGGGCTCCCAACGGCGCGCGCGCTGCTGCTGCTG 260  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysAspSerLeu 40  
 DB 261 TTCTGCTGCTGCTGCTGCGCGCACTATTGGAGAACCAAGTGGGGCCCTGCTCTTCC 320  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAlaCysAlaProMetGlyThrLeuLeuCysProSer 60  
 DB 321 TTTCCAGCGGAGAGAGGCTCTTCCGGGGCTTGGCCCCCAATGGGCTGCTCTTCCATCC 380  
 QY 61 SerAlaSerAlaLeuThrLeuLeuYThrSerArgCysAspAspAlaGlyThrGluLeuThr 80  
 DB 381 TCAAGCGTGAACCTGCTGCTTACACCAAGCGCTGACGAGATCGGGCACTGACGCTGACT 440  
 QY 81 GlyHisLeuValProHisIleAspGlyLeuArgValThrCysProGluSerGlyValHis 100

DB 441 GGCACCTGGTACCCACACAGATGGCTGAGGGTTTGGTGTCCAGAAATCCAGAGCCCAT 500  
 QY 11eProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyL1legly 120  
 DB 501 ATTCCCTTACACACAGCTCTCTAAGGCTGCCCCCTGAGCTGTGCGCTCTGAGGACATGGA 560  
 QY 121 GlyHisLeuSerProGluGlyLysLeuThrLeuProGluGluHisProCysLeuLysAla 140  
 DB 561 GGCACACCTTTCCCAACAGGGAGCTACACCTGCCGAGAGACACCGGTCTTAAGGCT 620  
 QY 141 ProAlaGluArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
 DB 621 CCAACGCTTCAAGATGCGACGCTCGCAAGCTGGACAGGGCCCCCGGCTCAGGGCAGGGGAA 680  
 QY 161 ArgSerProGluGluSerLeuGlyLysValArgAlaGlyLysArgAlaAsnThrAlaProGln 180  
 DB 681 AGGTCAACAGAAAGATCCCTGGGTGGGCTCGAAAGAAAGAAATGAATACACCCCCC 740  
 QY 181 PheGlnProProSerTyrglnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 741 TTCCAGCCCCCAGCTTACAGGCCACAGTGGCGGAGAACACAGCCAGAGGACCCCTGTT 800  
 QY 201 AlaserLeuArgAlaIleAspProAspGlyGlyGluAlaGlyArgLeuGlyThrThr 220  
 DB 801 GCATCCCTGAGGGCCATCGACCCCGACAGAGGTGAGGAGGTGCACTGGAGTACACATG 860  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
 DB 861 GATGCCCTCTTGAATAGCCGCTCCACACAGATTCCTCTCCGACCCAGTCACTGGTGA 920  
 QY 241 ValThrThrAlaGluGluLeuAspArgGlyLysThrLysSerThrHisValPheArgValThr 260  
 DB 921 GTAACCAACAGCGAGAGGTGATGCTGAGACCAAGACACCACTCTTACAGGCTACAG 980  
 QY 261 AlaGlnAspHisGlyIleCysProArgArgSerAlaLeuAlaThrLeuThrL1leuValThr 280  
 DB 981 GGGCAGAGACCAAGGATGCCCGAGAGATGCCCTGGCTTACCACTGATCTTGGTTACT 1040  
 QY 281 AspThrAspAspHisAspProValPheGluGluGlnGluTyrglyLysGlyLysLeuArgGlu 300  
 DB 1041 GACACCAATGACCAATGACCTCTGTTCAGACGACGAGATACAAAGAGAGCTCAGGAG 1100  
 QY 301 AsnLeuGluValGlyTyrglyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB 1101 AACCTGAGAGTTGGCTATGAGGTGCTCACTGTCAGGGCCACGAGATGATGCCCTCC 1160  
 QY 321 AsnAlaAsnL1leuTyrglyArgLeuLeuGlySerGlyLysProSerGluValPhe 340  
 DB 1161 AATGCCAATATCTGTATCCGCTGTGAGAGGGGTCTGGGGGACGCCCTCTGAAAGTCTTT 1220  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 DB 1221 GAGATCGACCTCGCTCTGCGGTGATCCGAAACCCGTGCGCTGTGATCGGAAAGAGTG 1280  
 QY 361 GluSerTyrglnLeuLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 DB 1281 GAATCTTACAGCTAGCGGTAGAGCAAGTGAACAGGGTCGGGACCCGGGCTCTCGAGT 1340  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 DB 1341 ACCACAGCGCGTGTTCCTTCTGTGTGAGAGATGACAAATGAATATGCCCCCAAGTTTAA 1400  
 QY 401 GluLysArgTyrglyValGluValArgGluAspValThrProGlyValaProValLeuArg 420  
 DB 1401 GAGAGCGGTATGTGTCCAGGTAGGAGATGAGATGCCAGGGGCCCATTAATCCGA 1460  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValValHisTyserL1leuSer 440  
 DB 1461 GTCAAGCCTTCGATCGAGACAGAGGGAGCAATGCCGTGTGTGCTACTATAGCATGAGT 1520  
 QY 441 GlyAsnAlaArgGlyGlyPheTyrglyLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 DB 1521 GGCATGCTCGGGAGACATTTTATCTGATGCCAGACTGAGAGCTGTGATGTGTGAGC 1580



OY	461	ProLeuAspTyrGluThrThrlsGlyTyrThrLeuArgValArgAlaAspGlyGly	480
Db	1581	CCCTCTGACTAAGACACGACCAAGAGATGACCTTACGGGTGCGACACAGATGTGTGGC	1640
OY	461	ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspTLeuAsnAsp	500
Db	1641	CGTCCCCCACTCTCAATGATCTCTGACTTGGAGACATGACAGGTCTTGATATTCACACAC	1700
OY	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu	520
Db	1701	AATGCCCCCATCTTGTCAGACACCCCTTCCAGGCTTACTGTCTTGAGAGATGTCCTTCA	1760
OY	521	GlyTyrTLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	1761	GGCTAACCTGGTTCCTCATGTCCAGGCTATCGACGCTGATGTGTGGACAAATGCCGCTCG	1820
OY	541	GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly	560
Db	1821	GAATACCCGCTTGGCTGTGGGTGGAGATGATCTTCCCTTCAACATCAACAAATGGCACAGGC	1880
OY	561	TrpIleSerValAlaAlaGluLeuAspArgGluGluValaAspPheTyrSerPheGlyVal	580
Db	1881	TGGATCTCTGTGGCTGTGAATCGACCGGAGGAAATTGATTTCTACAGCTTTGGGGTCA	1940
OY	581	GluAlaAspAspHisGlyTyrThrProAlaLeuThrAlaSerAlaSerValThrVal	600
Db	1941	GAAGCTGAGACCAATGGCATCTCCAGACATCACTGCTGGCGCAGTGCAGGTGATCTGTC	2000
OY	601	LeuAspValaAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620
Db	2001	CTGATGTGCAACGACACAAATCCAACTTTCACCAACAGATACACATGTCGGCTCAAT	2060
OY	621	GluAspAlaAlaValAlaGlyTyrSerValThrValSerAlaValaAspArgAspAlaHis	640
Db	2061	GAGGATGACAGTGTGGGCACCAAGCGTGTGACGGGTGACGTGGACCGATGATGCTCAT	2120
OY	641	SerValIleThrTyrGlnIleThrSerGlyLeuThrArgAsnArgPheSerIleThrSer	660
Db	2121	AGTGTATCATACCTTACCAATGATCAACAGTGGCAATCTGCAAAACCCCTTCTCATCAACAGC	2180
OY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGln	680
Db	2181	CAAGGTGTGGTGGCTGTATCCTTGGCCCTGGCACTGAGACTTGAACGGCGCAG	2240
OY	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleValVal	700
Db	2241	TATGTGTGGCTGTATCCGCTCCGATGGCACCTGGCGACGACACAGATGTGGTGTG	2300
OY	701	AsnValThrAspAlaAsnThrHisArgProAlaPheGlnSerSerHisTyrThrValAsn	720
Db	2301	AATGTACACGACGCCACACCCATGCTGTCTTTCACAGCTCCCACTAATACAGTAAT	2360
OY	721	ValaAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740
Db	2361	GTTATATGAGACCGGCGCGGACGACCAACGCTGGTGTGATCAACGCCACCGAATGAGAC	2420
OY	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2421	ACAGGTAGAAATGCCCGATCACTACTTCATGAGAGACACGATCCCCAGTTCCGCATC	2480
OY	761	AspAlaAspThrGlyAlaValaThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer	780
Db	2481	GATCAGACACACGGGGCGCTGTACCAACCGACTAGCTGACATATGAAACCAAGTGTCT	2540
OY	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnTyrSerAspThrThrTyr	800
Db	2541	TACACCCCTGGCCATTAATGCTCGGGACAAATGGCATTTCCCGAAGTCCGACACCACTTAC	2600
OY	801	LeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuAlaArgAspSerTyr	820
Db	2601	CTGAGATCTCTGTGAACGACGTAAATGACAAATGCCCTCTCAAGTTCTTCCGACAGATCTCTAC	2660

Qy	821	GlnGlySerValTYrgLyuAspValProPhePheThSerValLeuGlnIleSerIaThr	840
Db	2661	CAGGGCAGTGTCTATGAGGAGATGCCACCCTTCACTAGGCTTCGACGATCTCAGCCACT	2720
Qy	841	AspArgAspSerGlyLeuAsnGlyValArgValPheTYrThrPheGlnGlyValAspAspGly	860
Db	2721	GATGGTAATCTTGACCTTAATGGAAGGGCTCTTCTACACCTTCCAAAGAGGGGACGATGGA	2780
Qy	861	AspGlyAspPheIleValIleGlySerThrSerGlyIleValIArgThrLeuAspArgLeuAsp	880
Db	2781	GACGGTACCTTATTTGTTGAGTCCACGTCACAGGACATCGTGGAGAGCTTAACGAGGCTGGAT	2840
Qy	881	ArgGluAsnValAlaGlnTYrValLeuArgAlaTYrAlaValAspLysGlyMetProPro	900
Db	2841	CGAAGAAACGGGCGCCAGTATGCTTCCGGGCAATATCGATGAGCAAGGGGATGCCCA	2900
Qy	901	AlaArgThrProMetGlyValIleThrValIleValLeuAspValAsnAspAsnProProVal	920
Db	2901	GCCCGCAACCTTAAGAAAGTGAACGATCACTGTGTGATGTGAATGACAATCCCCCTGTC	2960
Qy	921	PheGlnGluAspGluPheAspValPheValIleGluAsnSerProIleGlyLeuAlaVal	940
Db	2961	TTTATGACGAGATAGAGTTTGAATGTGTTTGTGAAGAGAACAGCCCATTTGGGCTAGCCGTG	3020
Qy	941	AlaArgValIleThrAlaThrAspProAspGlnGlyIleThrAsnAlaGlnIleMetTYrGlnIle	960
Db	3021	GCCCGGGTCAACGCACTGACCCCGAAGAAAGCAACCATCCACAGATTATGTCACAGATT	3080
Qy	961	ValGlnGlyAsnIlePheProGlyValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla	980
Db	3081	GTGAGGGGCAACATCCTTGAGGCTCTTCACTGGACATCTTCTCCGGGAGGTGACAGCC	3140
Qy	981	LeuValAspLeuAspTYrGlyAspArgProGlyTYrValLeuValIleGlnAlaThrSer	1000
Db	3141	CTGGTAGACTTAAGCTACGAGAACCGGCGCTGAGTATGCTGTATCCAGGCAAGTCA	3200
Qy	1001	AlaProLeuValSerArgAlaIleThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	3201	GCTCTCTGTGTGAGCGGGCTACAGTCCAGTCCGCGCTCTTGAACGCAATGACACCCA	3260
Qy	1021	ProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTYrValIleThrAsnArgSerSerSer	1040
Db	3261	CCAGTGTGGGCAACTTGTAGATCCCTTTCAACAACATATGTCCATCCCTCAACAGC	3320
Qy	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	3321	TTCCCTCGGGGTCCTCATTTGGCCGAGTACCTGCCACATGACCTTATCTCAGTACTGCTG	3380
Qy	1061	ThrTYrSerPheGluArgGlyAsnGlyIleuSerLeuValLeuLeuAsnAlaSerThrGly	1080
Db	3381	ACTTACAGCTTTAGCGGGGAATGAACTCAGCTGTCTCTGTCATATGCTTCACAGGGT	3440
Qy	1081	GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal	1100
Db	3441	GACCTGAGCTTAAGCCGCGCACTGGAACAACAACCGGCTCTGTGAGGCTCATGAGACGTG	3500
Qy	1101	LeuValSer 1103	
Db	3501	CTGGTGTCA 3509	
RESULT 8			
LOCUS	AB065955	24370 bp	DNA linear PRI 23-JUL-2002
DEFINITION	Homo sapiens gene for seven transmembrane helix receptor, complete cds, isolate: CBRCTM_518.		
ACCESSION	AB065955		
VERSION	AB065955.1 GI:21929187		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		





```

QY 921 PheGluGlnAspGluPheAspValPheValGluGlnAsnSerProIleGlyLeuAlaVal 940
DB 2961 TTTGACGAGGATGATTGATTGATTGAGGAGAGAACAGCCCATTTGGGCTAGCCGTTG 3020
QY 941 AAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
DB 3021 GCCCGGGTCCACGACCTGACCCCGAGGAGGACCAATGCCAGATATGATCCAGATT 3080
QY 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
DB 3081 GTGGAGGGGCAATCCCTGAGGCTTTGACCTGAGCATCTTCTCCGGGAGGCTGACAGCC 3140
QY 981 LeuValAlaAspLeuAspTyrGluAspAsgProGluTyrValLeuValIleGlnAlaThrSer 1000
DB 3141 CTGGTAGACTTAACTACGAGGACCCGCTGAGTACGCTCTGATCATCCAGGCTAGCTCA 3200
QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020
DB 3201 GCTCTCTGTGAGCGGCGGTACAGTCCAGTCCCTCTTACCGCAATGACCAACCA 3260
QY 1021 ProValIleuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
DB 3261 CCAAGTCTGGGCAACTTTGAGATCTTTTCAACAATATGTCACCAATGCTCAACGAC 3320
QY 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
DB 3321 TTCCCTGGGGGTGTCATTGGCCGACGATCCCTGACCACTGATATCTCAGATAGTCTG 3380
QY 1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
DB 3381 ACTTACAGCTTTGAGGGGAGAAATGAACTCAGCTGCTGCTGCTCAATGCTTCAACGAG 3440
QY 1081 GluLeuAlaLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
DB 3441 GAGCTAGAGTAACTGCGGACGACGACGACCAACACGCGCTCTGAGGCTCATGACGCTG 3500
QY 1101 LeuValSer 1103
DB 3501 CTGGTGTCA 3509

```

```

RESULT 9
AL390252 169241 bp DNA linear PRI 08-JUN-2001
LOCUS Human DNA sequence from clone RP11-29704 on chromosome 1, complete
DEFINITION
ACCESSION AL390252
VERSION AL390252.9 GI:14346077
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169241)
REFERENCE
AUTHORS Thomas D.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

```

On Jun 11, 2001 this sequence version replaced gi:13273808.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; Sw.,  
 SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Info, information on the WormPeP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contig of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>  
 RP11-29704 is from the library RPCI-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6  
 This sequence is the entire insert of clone RP11-29704 The true  
 left end of clone RP11-173K24 is at 98796 in this sequence. The  
 true right end of clone RP11-352P4 is at 101180 in this sequence.

## FEATURES

## SOURCE

```

1..169241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_id="RP11-29704"
/location="RPCI-11.2"
1676..1775
/note="MIR repeat: matches 45. 152 of consensus"
2574..3410
/note="CpG island"
/evidence="not_experimental"
2809..2936
/note="64 copies 2 mer gg 60% conserved"
2814..2943
/note="13 copies 10 mer ggcggggcgg 63% conserved"
2982..2987
/note="sequence from overlapping clones B352P4 (
AL356389) and dJ641D22 (AL157901). Assembly confirmed by
restriction digest."
3007..3116
/note="11 copies 10 mer gcccgggggc 63% conserved"
3364..3411
/note="16 copies 3 mer tgc 83% conserved"
9894..10207
/note="AluSg repeat: matches 1. 296 of consensus"
10332..10617
/note="AluJ repeat: matches 10. 299 of consensus"
11683..12328
/note="CpG island"
/evidence="not_experimental"
12551..12727
/note="MIR repeat: matches 10. 197 of consensus"
13355..13442
/note="44 copies 2 mer gg 65% conserved"
15405..15448
/note="22 copies 2 mer gt 75% conserved"
17917..17995
/note="MIR repeat: matches 45. 129 of consensus"
19254..19293
/note="4 copies 10 mer aggcctgggg 95% conserved"
19651..19956
/note="AluSg repeat: matches 5. 309 of consensus"
19972..20270
/note="AluSc repeat: matches 1. 296 of consensus"
26166..26195
/note="10 copies 3 mer gag 90% conserved"
29612..29726
/note="FLAM C repeat: matches 1. 118 of consensus"
29778..30077
/note="AluX repeat: matches 1. 298 of consensus"
30135..30428
/note="AluY repeat: matches 1. 298 of consensus"
30454..30616
/note="MIR repeat: matches 98. 256 of consensus"
32159..32470
/note="AluSg repeat: matches 1. 312 of consensus"

```

repeat\_region 32481..32780  
 /note="AluX repeat: matches 1..298 of consensus"  
 repeat\_region 34670..34803  
 /note="MIR repeat: matches 123..250 of consensus"  
 misc\_feature 36339..36835  
 /note="CpG island"  
 /evidence=not\_experimental  
 repeat\_region 37313..37937  
 /note="MER41B repeat: matches 1..635 of consensus"  
 repeat\_region 37938..38051  
 /note="MER41B repeat: matches 103..213 of consensus"  
 repeat\_region 38052..38351  
 /note="AluY repeat: matches 1..300 of consensus"  
 repeat\_region 38352..38753  
 /note="MER41B repeat: matches 213..635 of consensus"  
 repeat\_region 38987..39182  
 /note="MER41B repeat: matches 1..171 of consensus"  
 repeat\_region 39183..39482  
 /note="AluX repeat: matches 1..300 of consensus"  
 repeat\_region 39483..39504  
 /note="MER41B repeat: matches 170..191 of consensus"  
 repeat\_region 39537..39824  
 /note="AluX repeat: matches 1..304 of consensus"  
 repeat\_region 39850..40383  
 /note="MER41A repeat: matches 1..554 of consensus"  
 repeat\_region 40394..40483  
 /note="MER41D repeat: matches 359..448 of consensus"  
 repeat\_region 40532..40838  
 /note="AluY repeat: matches 1..311 of consensus"  
 repeat\_region 40956..41169  
 /note="L1P81 repeat: matches 5927..6151 of consensus"  
 repeat\_region 41170..41472  
 /note="MER41D repeat: matches 433..745 of consensus"  
 repeat\_region 41473..41771  
 /note="AluY repeat: matches 2..300 of consensus"  
 repeat\_region 41772..42011  
 /note="MER41D repeat: matches 745..981 of consensus"  
 repeat\_region 42399..42702  
 /note="AluX repeat: matches 1..304 of consensus"  
 repeat\_region 42948..43256  
 /note="AluY repeat: matches 1..310 of consensus"  
 repeat\_region 44024..44141  
 /note="MIR repeat: matches 21..147 of consensus"  
 repeat\_region 44187..44622  
 /note="MIR repeat: matches 77..547 of consensus"  
 repeat\_region 44790..45097  
 /note="AluX repeat: matches 1..306 of consensus"  
 repeat\_region 45098..45232  
 /note="AluSg/x repeat: matches 1..136 of consensus"  
 repeat\_region 46712..47013  
 /note="AluY repeat: matches 1..305 of consensus"  
 repeat\_region 47032..47323  
 /note="AluX repeat: matches 1..292 of consensus"  
 repeat\_region 47669..47826  
 /note="L2 repeat: matches 2545..2709 of consensus"  
 repeat\_region 48017..48204  
 /note="MIR repeat: matches 22..209 of consensus"  
 repeat\_region 49692..49988  
 /note="AluX repeat: matches 1..297 of consensus"  
 repeat\_region 54709..55019  
 /note="AluSg repeat: matches 1..311 of consensus"  
 repeat\_region 55261..55380  
 /note="60 copies 2 mer gt 79% conserved"  
 repeat\_region 55267..55386  
 /note="12 copies 10 mer gtgtgtgtgt 79% conserved"  
 repeat\_region 55386..55419  
 /note="17 copies 2 mer gt 88% conserved"  
 repeat\_region 55756..56029  
 /note="L2 repeat: matches 46..328 of consensus"  
 repeat\_region 56258..56327  
 /note="MIR repeat: matches 323..388 of consensus"  
 repeat\_region 56425..56915  
 /note="L2 repeat: matches 330..834 of consensus"

repeat\_region 56926..57541  
 /note="L2 repeat: matches 2097..2748 of consensus"  
 repeat\_region 58447..58523  
 /note="MIR repeat: matches 49..146 of consensus"  
 repeat\_region 59152..59300  
 /note="MIR repeat: matches 29..185 of consensus"  
 repeat\_region 60773..61014  
 /note="MIR repeat: matches 4..262 of consensus"  
 repeat\_region 61115..61227  
 /note="MIR repeat: matches 36..150 of consensus"  
 repeat\_region 61443..61519  
 /note="L1M1 repeat: matches 6234..6314 of consensus"  
 repeat\_region 61956..62063  
 /note="MER5A repeat: matches 35..141 of consensus"  
 repeat\_region 63402..63437  
 /note="18 copies 2 mer ta 80% conserved"  
 repeat\_region 68999..69310  
 /note="AluB repeat: matches 2..306 of consensus"  
 repeat\_region 69476..69781  
 /note="AluX repeat: matches 1..292 of consensus"  
 repeat\_region 70577..70718  
 /note="MER3 repeat: matches 2..155 of consensus"  
 repeat\_region 71406..71699

Alignment Scores:  
 Pred. No.: 0  
 Score: 1103.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 37.74%  
 DB: 9  
 Gaps: 0

US-09-916-849a-3 (1-2923) x AL390252 (1-169241)

Qy	1	MeRargserProAlaThrGlyValProLeuProThrProProProLeuLeuLeu	20
Db	3330	ATGGGAGCCCGGACCGAGGCTCCCTCCCAACGCGCGCGCTGCTGCTG	3389
Qy	21	LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu	40
Db	3390	TTGGTGTGCTGCGTGGCGCGCGCACCTTATGGAGACCAATGGGCGCTTCGTTCTTG	3449
Qy	41	GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysProSer	60
Db	3450	GGGTCGAGGAGGAGGAGGCTCTGCGGGGCTGGCGCCCTCAATGGCTGCTGCTGCTG	3509
Qy	61	SerAlaSerLeuLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr	80
Db	3510	TCAAGCTGAACCTCTGCTCTTACACAGCGCTGACGAGATGGGCACTGAGCTGACT	3569
Qy	81	GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis	100
Db	3570	GGCAGCTGTATCCCAACCAACAGATGGCTTGGATGGTTCAGATTCAGAGGCCAT	3629
Qy	101	HisProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyHis	120
Db	3630	ATTCCCTTACACACAGCTCTGTAAGGCTGCCCTGGAGCTGCTGCTGCTGCTGCTG	3689
Qy	121	GlyHisLeuSerProGlnGlnGlyValLeuThrLeuProGlnGlnHisProCysLeuValAla	140
Db	3690	GGCCACCTTCCCAACGAGGAGGAGCTCACTGCGGAGGAGGACCCGCTTAAAGGCT	3749
Qy	141	ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu	160
Db	3750	CCAGGCTCAGATGCCAGTCTCTCAAGTGTGACAGGCCGCCGAGGAGGAGGAGGAG	3809
Qy	161	ArgSerProGlnGlnSerLeuGlyGlyArgValArgAsnValAsnThrAlaProGln	180
Db	3810	AGGTCAACAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3869
Qy	181	PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal	200
Db	3870	TTCCAGCCGCCCACTTCCAGGACCAAGTCCGAGAACCAAGCAAGGACCCCTGTT	3929

QY 201 AIASeLeuAArgAlaIleaspProaPgiugIyGluAlaGlyArgLeuGluTyThrMet 220  
DB 3930 GGATCCCTGAGGGGCGCATCCACCGGAGGAGGAGGAGCTGAGCTGAGATGACCATG 3989  
QY 222 AAPAlaLeuPheAspSerArgSerAasnGlnphePheSerLeuAspProValThrGlyAla 240  
DB 3990 GATGCCCTCTTGTGATGCGGCTCAACAGTTCTTCTCCGTGAGCCAGTCACTGGTGA 4049  
QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
DB 4050 GTTACCAAGCCGAGAGAGCTGATCTGTGAGACCAAGACACCACTCTTCAGGGTCAAG 4109  
QY 261 AIAgInaPheHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 4110 GCGCAGGAGCACGGCGATGCCCGACCAAGATGCTGCTGCTACACTCACTCTGGTTACT 4169  
QY 281 AAPThrAsnAspHisAspProValPheGluGlnGluTyThrLysSerLeuAspArgLys 300  
DB 4170 GATCCCAATGACCATGACCTGTGTGAGCAGCAGAGATCAAGAGAGCTCAAGGGAG 4229  
QY 301 AsnLeuGluValGlyTyArgLysValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 4230 AACCTGAGAGGTGGCTATGAGGTCTCACTGTCAAGGCCACCGATGTGTATGCCCTCCC 4289  
QY 321 AsnAlaAsnIleLeuTyArgLeuLeuGluGlySerGlyLysProSerGluValPhe 340  
DB 4290 AATGCCAATATCTGTACCGCGCTGCGAGGGGTCTGGGGGAGCGCCCTGAGACTTT 4349  
QY 341 GJuIIeaSPProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
DB 4350 GAGATGACCTGTGCTGTGGGTGATCCAGCCCGTGGCCCTGTGTATGGGAGAGGGTG 4409  
QY 361 GJuSerTyArgLysLeuThrValGluAspAspGlnGlyArgAspProGlyProArgSer 380  
DB 4410 GATCTTACCACTGACGCTGATGAGGCACTGACAGGGTGGGAGCCCGGCTCTGGAGT 4469  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
DB 4470 ACCACAGCGCGTGTCTTCTTCTGTGAGATGATGATGATGATGATGATGATGATGAT 4529  
QY 401 GJuLysArgTyValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
DB 4530 GAGAAAGCTGATGTGCTCAAGGTGAGGAGAGATGATGATGATGATGATGATGATGAT 4589  
QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValIleIleTySerIleMetSer 440  
DB 4590 GTCAAGCGCTCGGATCGAGACAGAGGGAGCAATGCGTGTGCTACATATGATCATGAGT 4649  
QY 441 GlyAsnAlaArgGlyGlnPheTyThrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
DB 4650 GGCATATGCTCGGGGACAGATTATCTGAGATGCCAGACTGAGCTCTGATGTGTGAGC 4709  
QY 461 ProLeuAspTyArgLysThrLysGluTyThrLysAspValArgAlaGlnAspGlyGly 480  
DB 4710 CCTCTTGACTATGAGACGACCAAGAGTACACCTTACGGGTGAGAGCAGACAGATGGTGC 4769  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 4770 CGTCCCGCATCTCTTAATGTCTGTGCTGTGAGACGTACAGTCTGATATTCACAGAC 4829  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
DB 4830 AATGCCCGCATCTTGTGTGAGACCCCTTTCAGGCTACGTCTGTGAGGTCTCCCTTA 4889  
QY 521 GlyTyThrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
DB 4890 GGTACTCTGTCTCTCATGTCCAGGCTATCGACGCTGATGTGTGATCATATCCCGCTG 4949  
QY 541 GJuTyArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyThrGly 560  
DB 4950 GAATACCGCTGTGTGTGGGTGGAGCATGACTTCCCTTCAACATATGGACAGGCG 5009

QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTySerPheGlyVal 580  
DB 5010 TGGATCTCTGTGGCTGTGAACTGGACCGGAGAGAAATTGATTCTACGCTTTGGGGTA 5069  
QY 581 GJuIleAspAspHisGlyTyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
DB 5070 GAGCTCGAGACCATGGCATGCTCAGCACTGCTGTGGCAGGTGACGCTGTCTGC 5129  
QY 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsn 620  
DB 5130 CTGATGTCAACGACCAACATTCACCTTACCAACAGATGACACAGTGGCTCAT 5189  
QY 621 GJuAspAlaAlaValGlyTyThrSerValThrValSerAlaValAspArgAspAlaHis 640  
DB 5190 GAGGATGACGCTGTGGGACCAAGCGTGTGACGGGTGACGCTGTGACCGTGAATGCTAT 5249  
QY 641 SerValIleThrTyArgLysIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
DB 5250 AGTGTATCACTTACCAATGACAGTGGCAATCTGAAACCGCTTCTCATATCACAGC 5309  
QY 661 GlnSerGlyGlyTyLeuValSerLeuAlaLeuProLeuAspTyTyLysGluValArgGln 680  
DB 5310 CAAAGTGTGTGGGCTGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5369  
QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
DB 5370 TATGTGTGTGTGTATCGCTCCAGTGGCACTGGCGAGCACGGGACAGATGTGTGTG 5429  
QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyThrValAsn 720  
DB 5430 AATGTACCAAGCGCCACCAACCATGTGCTGTCTTCAAGACTCCCATATACATGAT 5489  
QY 721 ValAsnGluAspArgProAlaGlyTyThrValValLeuIleSerAlaThrAspGluAsp 740  
DB 5490 GTTATATGAGACCGGCGGAGGACGACCAAGGTGTGTGATCAAGCGCACGAGTGAAGAC 5549  
QY 741 ThrGlyGluAsnAlaArgIleThrTyPheMetGluAspSerIleProGlnPheArgIle 760  
DB 5550 ACAGGTGAGATGCGCGATCACTTCAATGAGAGACACATGCCCTCCAGATTCGGATTC 5609  
QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyArgLysAspGlnValSer 780  
DB 5610 GATGACAGACCGGGGCTGTACCAACCGCTGAGCTGAGCTATGAAAGCAATGTCT 5669  
QY 781 TyThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTy 800  
DB 5670 TACACCTGGCATTAATGCTCGGAGACATGCAATGCCCAAGATCCGACACCACTTAC 5729  
QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgSerTy 820  
DB 5730 CTGAGATCTCTGTGAAACGAGTGAATGACATGCCCTCAGTTCCTGCGAGACTCTTAC 5789  
QY 821 GlnGlySerValTyArgLysAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
DB 5790 CAGGGCAGTGTCTATGAGAGATGTCCACCTTCACTAGCGTCTGCAATCTCAGCCACT 5849  
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyLysAspAspGly 860  
DB 5850 GATGTGATTTCTGACTTAATGAGCAGGCTTTCACACTTCAAGAGGAGGAGAGATGA 5909  
QY 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880  
DB 5910 GACGTGACTTATATGTGAGTCCACGTGAGGACATGCGACCAAGCTACGAGGCTGTAT 5969  
QY 881 ArgGluAsnValAlaGlnTyThrValLeuAspArgAlaTyAlaValAspLysGlyMetProPro 900  
DB 5970 CGAGAGAAAGTGGCCAGTATGTCTTGGCGCAATGCAAGTGAAGGAGATGCCCA 6029  
QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920  
DB 6030 GCCCGCACACCTTATGAGAGTGAAGTCACTGTGTGTGATGATGATGATGATGATGATG 6089  
QY 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940



[illegible]

	Insert size: 227194; sum-of-contigs
	Insert size: 168515; 3.2% error; agarose-1p
	Quality coverage: 7.72x in Q20 bases; sum-of-contigs Quality
	coverage: 10.52x in Q20 bases; agarose-1p
	-----
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 1 contigs. Gaps between the contigs
	* are represented as runs of N. The order of the pieces
	* is believed to be correct as given, however the sizes
	* of the gaps between them are based on estimates that have
	* provided by the submittor.
	* This sequence will be replaced
	* by the finished sequence as soon as it is available and
	* the accession number will be preserved.
	1 227194: contig of 227194 bp in length.
FEATURES	
Source	Location/Qualifiers
	1..227194
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP11-6121"
	/clone_1fb="RPC1-11.1"
	1..227194
	/note="assembly_fragment:02685"
	clone_end:SP6
	vector_side:right
	clone_end:T7
	vector_side:left"
BASE COUNT	64124 a 52119 c 51230 g 59721 t
ORIGIN	
Alignment Scores:	
Prod. No.:	0 Length: 227194
Score:	1103.00 Matches: 1103
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	37.74% Indels: 0
DB:	2 Gaps: 0
US-09-916-849A-3 (1-2923) x BX284647 (1-227194)	
QY	1 MetAaSerGProAlaThrGlyValProLeuProThProProProProLeuLeuLeu 20
Db	3330 ATGGGAGCCCGGCGCAACGGCGTCCCTCCCAACGCGCGCGCGCTGCTGCTG 3389
QY	21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyAAspGlnValGlyProCysArgSerLeu 40
Db	3390 TTGCTGCTGCTGCTGCGCGCGCACTTATGGAGACCAATGGGGCCCTGTCCTTCTTG 3449
QY	41 GlySerArgGlyArgGlySerSerGlyValAAspAlaProMetGlyTyrPLeuCysProSer 60
Db	3450 GGGGTCCAGGAGAGAGGCTCTTGCGGGGCGCTGCGCCCAATGAGGCTGCTGCTCAATCC 3509
QY	61 SerAlaSerLeuLeuTyrPLeuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80
Db	3510 TCAAGCTTCGAACCTCTGAGCTCTTCAACCAAGCGCTGACGAGATGCGGACCTAGACTACT 3569
QY	81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100
Db	3570 GGGCACTGGTATCCCAACCAACGATGGCTGAGGGGTTGGTGTCCAGATCCGAGGCCCAT 3629
QY	101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyTyrLeu 120
Db	3630 ATTTCCTTACACAGCTCTCTGAAGGTGCCCCCTGAGAGCTGTGCTGCGGCAATTGGA 3689
QY	121 GlyHisLeuSerProGlnGlyTyrLeuThrLeuProGlnGlnHisProCysLeuTyrAla 140
Db	3690 GGGCACTTTCCCAACAGGCAAGCTCAACATCCCGAGGAGCAACCCGTCTTAAAGCT 3749
QY	141 ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160
Db	3750 CCAAGGCTCAGAGTCCAGTCTCTGCAAGCTGGCAACGAGCCCGGCTCAAGGCAAGGGGAA 3809



QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgLysArgAsnValAsnThrAlaProGln 180  
 Db AGGTGACAGAGAGAGTCCCTGGGTGGCGCTCGGAAAAGAAATTAAATCAGCCCCCAG 3869  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db TTCCAGCCCCCGCTGACAGGCGCACAGTCCGGAACCCAGCCAGCAGGACCCCTGTT 3929  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyArgLeuGlyLysThrMet 220  
 Db GATTCCTGAGGAGCCATCGACCCGAGAGGAGGAGGAGGAGTGACTGAGTGAACACAG 3989  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db GATGCCCTCTTGAATAGCGCGCTCCACACAGTTCTTCCCTGGACCCAGTCACTGGTGA 4049  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
 Db GTAAACCAAGCCGAGGAGCTGATCGTGAACCAAGACCAACCCCTCTTCAAGGGTCAAG 4109  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db GGCGAGGACCAAGCCAGTCCGACAGAGTGGCTGCTACCTCACTCACTCACTTGTACT 4169  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlyLysArgLysSerLeuArgGlu 300  
 Db GACACCAATGACATGACCTGTGTTGAGACAGACAGAGTACAGAGAGCTCAAGGAG 4229  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db AACCTGAGAGTGGCTATGAGGTCTCACTGTCAAGGCGCACGATGATATGCCCCCTCC 4289  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysSerProSerGluValPhe 340  
 Db AATGCCAATTTCTGATCCGCTGCTGAGAGGCTCTGGGGCAGCCCCCTGTAGTCTTT 4349  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
 Db GAGATGACCCCTGCTGCTGGGTGATCCGAAACCGTGGCTGTGATCGGAGAGGAG 4409  
 QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProAspSer 380  
 Db GAATTCCTACAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4469  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
 Db ACCAAGCCGCTGTTCTTCTGCTGAGGAGTGAATGATGATGATGATGATGATGATGAT 4529  
 QY 401 GluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db GAGAAACGCTATGTGCTCAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4589  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValIleHisTyrSerIleMetSer 440  
 Db GTACAGAGCTCGGATGAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4649  
 QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValIleSer 460  
 Db GGCAATGCTCGGGGAGGAGTTTATCTGATGCCAGACTCGAGCTCTGATGAGTGAAGC 4709  
 QY 461 ProLeuAspTyrGluThrThrLysGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
 Db CCTCTGACATATGAGACAGACAGAGGAGTACACCTTCAGGGTGCAGAGACAGAGTGGC 4769  
 QY 481 ArgProProLeuSerAsnValIleSerGlyLeuValThrValGlnValLeuAspIleAsnArg 500  
 Db CGTCCGCCATCTCTAATGCTCTGAGCTTGTGATCAAGTACAGGCTCTGATACAGAC 4829  
 QY 501 AsnAlaProIlePheValIleThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db AATGCCCCATCTTCTGTCAGACACCCCTTCCAGGCTACTGCTCTGAGAGAGTGTCCCTTA 4889

QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db GGCTACCTGGTTCTCCATGTCAGGCTATCGACGCTGATGCTGATGATGATGATGATGAT 4949  
 QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db GAATACCCCTGCTGGGTGGGAGCATGATCTTCCCTTCAACATCAATGAGCAGGCGC 5009  
 QY 561 TrpIleSerValAlaAlaGlyLeuAspArgGluValAspPheTyrSerPheGlyVal 580  
 Db TGGATCTCTGTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5069  
 QY 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db GAAGCTCGAGACCATGGACCTCCAGACCTCACTGCTGGCCAGGTGAGCTGACTGTC 5129  
 QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 Db CTGATGTCAAGAACAAATCAACCTTACCCAAACAGAGTACAGAGTGGGCTCAAT 5189  
 QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 Db GAGATGACAGCTGTGGGACCAAGCTGTGAGGAGTGTGATGATGATGATGATGATGAT 5249  
 QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db AGTGCATCACTTACCATGATCAACAGTGGCAATATCCGAACCGCTTCTCCATCAACAGC 5309  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArgGln 680  
 Db CAAAGTGGGTGGGCTGGTATCCCTGCTGCGCACTGAGTACAACTTGAAGGCGCAG 5369  
 QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db TATGCTGTGGCTTACCGCTCCGATGGCACTCGGAGGAGACCGGACAGTTGGTG 5429  
 QY 701 AsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 Db AATGTCAACGACCGCAACACCATGCTGTCTTCAAGAGCTCCCACTATCACTGAT 5489  
 QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db GTTAAATGAGACCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5549  
 QY 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 Db AAGGTGAGAAATGCCGATCATCTTCAATGAGGAGCAGCATCCCCAGTTCCGCAATC 5609  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db GATGAGACACCGGGGCTGTCCACCAAGCTGAGCTGAGCTATGAGAACCAAGTGTCT 5669  
 QY 781 TyrThrLeuAlaIleThrAlaArgAsnGlyIleProGlnLysSerAspThrThrTyr 800  
 Db TACACCTCGGCATTAATGCTCTGGGACATGGCATTTCCCGAAGTCCGACACCTTAC 5729  
 QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db CTGAGATCTCGTGTAACAGCTGATGACATGACCTTCACTTCTGAGGAGCTCTTAC 5789  
 QY 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db CAGGAGAGTGTATGAGAGATGTGACCCCTTCACTAAGGTCTTGAGATGTCAAGCAT 5849  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db GATCGTGAATCTTGACATTAATGAGGAGGCTTCTTACACCTTCAAGAGGAGGAGCATG 5909  
 QY 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 Db GACGGTGACTTATATGATGAGTCCAGTCAAGGACATGTGCGAACCTACGAGGCTGAT 5969  
 QY 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900

```

Db      5970 CGAAGAAACGCGCCAGTATGTCTTGGGGGCAATTCAGTGAACAGGGAGTGGCCCCA 6029
Qy      901 AAlaArgThrProMetGluVal1ThrVal1ThrVal1LeuAspVal1AsnAspAsnProProVal1 920
Db      6030 GCCCGCACACCTATGAAAGTACAGTCACTGTGTTGATGTGAATGACAAATCCCTGTC 6089
Qy      921 PheGluGluAspGluPheAspVal1PheVal1GluGluAsnSerProGluLeuVal1 940
Db      6090 TTTAGCAGGAGGAGTGTGATGTGTGTGTGAGAGAACACCCCATGGGCTGAGCCGTG 6149
Qy      941 AAlaArgVal1ThrAlaThrAspProAspGluGlu1ThrAsnAlaGln1MetTyrGlu1Ile 960
Db      6150 GCCCGGCTCACAGCACACTGACCCGATTAAGCACAATGCCCATGATTATATACAGATT 6209
Qy      961 ValGluGluAsn1IleProGluVal1PheGluLeuAsp1IlePheSerGluGluLeuThrAla 980
Db      6210 GTGAGAGGCAACATCCCTGAGGCTTTTCAGCTGAGCATCTTCGGGGAGCTGACAGCC 6269
Qy      981 LeuVal1AspLeuAsp1TyrGluAspArgProGluTyrVal1LeuVal11IleGln1ThrSer 1000
Db      6270 CTGGTGAACCTTGAAGCTACGAGACCGGCTGAGTACGTCCTGTCATCCAGCCACGTC 6329
Qy      1001 AAlaProLeuVal1SerArgAlaThrVal1HisVal1ArgLeuAspArgAsnAspAsnPro 1020
Db      6330 GCTCTCTGTGAGACCGGCTACAGTCACTCCGCTCTTGACCGCATGACACCA 6389
Qy      1021 ProVal1LeuGluAsn1PheGlu1IleLeuPheAsnAsnTyrVal1TrrAsnArgSerSer 1040
Db      6390 CCAATGCTGGGCACTTGAATCCTTTTCAACAACATATGTCACATCCGTCAGAGCAGC 6449
Qy      1041 PheProGluGluVal1IleGluVal1ProAlaHisAspProAsp1IleSerAspSerLeu 1060
Db      6450 TTCCCTGGGGTGCATTGGCCGAGTACCTGCCATGACCTGATATCTCAGATATGCTG 6509
Qy      1061 ThrTyrSerPheGluArg1AsnGluLeuSerLeuVal1LeuAsnAlaSerThrGly 1080
Db      6510 ACTTACGCTTGAAGCGGGAAATGAATCACTGCTGCTCTCAATGCTCCACGGGT 6569
Qy      1081 GlnLeuLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluVal1IleMetSerVal1 1100
Db      6570 GAGCTGAAGCTAACCCGCGACTGACACAAACCGGCTCTGAGGCGCATATAGGCTG 6629
Qy      1101 LeuVal1Ser 1103
Db      6630 CTGGTGTCA 6638

```

```

RESULT 11
AL157901/c      116603 bp  DNA  linear  HTG 10-JUL-2001
LOCUS DEFINITION Homo sapiens chromosome 1 clone RP4-641D22 map pl3.1-13.3, ***
SEQUENCING IN PROGRESS ***, 9 unordered pieces.
AL157901
ACCESSION AL157901.5 GI:9796614
VERSION HTG: HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS McLay, K.
TITLE Direct Submision
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9212479.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d0641D22

```

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator ABI; 3% of reads
Chemistry: Dye-terminator Big Dye; 96% of reads
Consensus quality: 112460 bases at least Q40
Consensus quality: 114130 bases at least Q30
Consensus quality: 115038 bases at least Q20
Insert size: 115803; sum-of-contigs
Insert size: 115316; 1.9% error; agarose-fp
Quality coverage: 3.82x in Q20 bases; sum-of-contigs Quality
coverage: 4.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3348: contig of 3348 bp in length
* 3349 3348: gap of 100 bp
* 3449 13632: contig of 10184 bp in length
* 13633 13732: gap of 100 bp
* 13733 20620: contig of 6888 bp in length
* 20621 20720: gap of 100 bp
* 20721 24173: contig of 3453 bp in length
* 24174 24273: gap of 100 bp
* 24274 41397: contig of 17124 bp in length
* 41398 41497: gap of 100 bp
* 41498 81263: contig of 39766 bp in length
* 81264 81363: gap of 100 bp
* 81364 92945: contig of 11582 bp in length
* 92946 93045: gap of 100 bp
* 93046 97949: gap of 4904 bp in length
* 97950 98049: gap of 100 bp
* 98050 116603: contig of 18554 bp in length.
Location/Qualifiers
1. 116603
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p13.1-13.3"
/clone="RP4-641D22"
/clone_lib="RPCI-4"
1. 3348
/note="assembly_fragment:00133
fragment_chain:1"
3449. 13632
/note="assembly_fragment:01128
fragment_chain:1"
13733. 20620
/note="assembly_fragment:00187
fragment_chain:2"
20721. 24173
/note="assembly_fragment:00757
fragment_chain:2"
24274. 41397
/note="assembly_fragment:00109"
41498. 81263
/note="assembly_fragment:00393"
81364. 92945
/note="assembly_fragment:00527"
93046. 97949
/note="assembly_fragment:00530"
98050. 116603
/note="assembly_fragment:00873"
29209 a 29209 c 28975 g 28294 t 800 others

```

```

BASE COUNT 29325 a 29209 c 28975 g 28294 t 800 others
ORIGIN
Alignment Scores:

```

Pred. No.:	0	Length:	116603
Score:	1091.00	Matches:	1091
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.32%	Indels:	0
DB:	2	Gaps:	0

US-09-916-849A-3 (1-2923) x AL157901 (1-116603)

[illegible]

QY	333	GLVGLYSERPROSERGIUVALPHEGIIULEAPPROARGSERGIYVALILAEGRHARG	352
Db	68596	GGGGGGACGCCCTCTGAAGCTTTGAAGTGAACCCCTCGCTGGGGGTGATCCGAACCCCT	68537
QY	353	GIYPROVALASPARAGIUGIULVALIGUSERYRGIIULETHYRVALGIUNLASERASPGIN	372
Db	68536	GGCCCTGTGGATCGGGAGAGGTGGAAATCTTACAGCTACACGGTAGAGGCMAGTGACCAAG	68477
QY	373	GIYARGASPROGLYPROARGSERTHYRVALAVALPHELEUSERVALGIUNSEAPAP	392
Db	68476	GGTCGGGACCCGGGGTCTCGAGATACCAAGCCGGCTTTCTCTGTGAGGAYTGAC	68412
QY	393	ASNAASPMALAPROGINPHESERGIULYSARGYVALVALGINVALARGGIUNSEAPVAL	412
Db	68416	AATGATTAATCCCCCAAGTTTATGTAGAAACCGCTATGTGTCCAGGTAGAGAGATGAG	68355
QY	413	THYRPROGIYALAPROVALLEUARGVALTHYRILASERASPARASPIYSGIYSEASNALA	432
Db	68356	ACTCCAGAGGGCCCAAGTACCTCCAGTCCAGACGCTCGAGTCGAGACAGAGGGAGCATGCC	68297
QY	433	VALVALHISERYSERIILEMETSERGIYASRGYVALVALGINVALPHELYRLEUASPMALGIN	452
Db	68296	GTGGTGGACCTATAGACATCATGATGGGAGATGCTGGGGACAAGTTTATCTGGATGCCAG	68237
QY	453	THIRGIYALALEUASPMYVALISERPROLEUASPYRGIIUTHYRTHYRISYGLIYRTHLEU	472
Db	68236	ACTGAGACTCTGGATGTGGTGAAGCCCTCTTGAATTAGAGACGCAAGGAGTACACCTTA	68177
QY	473	ARGYVALARGALAGIINSPGIYGIYARGPROPROLEUSERASPMYVALSERGIYLEUVALTHYR	492
Db	68176	CGGGTGGACACACAGATGGGTGGCCGTCCCACTCTCAATGTCCTGGCTTGGTGGACA	68117
QY	493	VALGINVALLEUASPMIILEASPMASPMALAPROILEPHEVALSERTHYRPROHEGINALA	512
Db	68116	GTAACAGTCTCTGATATTCACACGACATGCCCCCATTTGTGTACAGACCCCTTTCCAGGCT	68055
QY	513	THYRVALLEUGIUNSERVALPROLEUGIYTYRLEUVALLEUHIIVYALGINALILEASPMALA	532
Db	68056	ACTGTCTCTGAGAGCGTCCCTTAAAGCTACCTGTGTCTCCATGTCACAGGCTATTCAGCGCT	67997
QY	533	ASPMALAGIYASPMASPMALAAARGLEUGIUYTYRARGLEUALAGIYVALGIYHISAPPHERO	552
Db	67996	GATCTGTGTATCAATGCCGCCCTCGAATAACCGCTTGCTGGGTGGGACATGACTTCCCC	67933
QY	553	PHETHYRILEASPMANGIYTHNGIYTPRIILESERVALALAGIILEUASPMARGIUGIUN	572
Db	67936	TTTCAACATCAACAATGGCAACAGCTGTGATCTGTGGCTGTGAATGTGACCGGGAGGAAA	67877
QY	573	VALASPMHETRYSERPHEGIYVALGIUALAARGASPMHIGIYTHYRPROILAEUTHYRALA	592
Db	67876	GTTATATTTTCAACGCTTGGGGGTGAAGGCTCGAGACCAATGGCACTCAGACATCACTGCC	67817
QY	593	SERILASERVALSERVALTHYRVALLEUASPMVALASPMASPMANSPROTHYRPHETHYRGIN	612
Db	67816	TGGCCAGGTGTACAGCGGTGACTGTGCTGGAGTCAACAGACATTCMAACTTTATCCCAA	67755
QY	613	PROGIUYRTHYRVALARGLEUASPMANGIUNSPMALAVALAGIYTHYRSEYVALIYTHYRVAL	632
Db	67756	CCAGAGTACACAGTGGCGCTCAATGAGAGAGCGAGCTGTGGCCACAGCGTGTGGAGCGGAG	67697
QY	633	SERILAVALLASPARASPMALAHISERVALILETHYRGTGINILETHYRSEYGIYASPMTHYR	652
Db	67696	TCACTGTGTGACCTGTATGCTCATAGTGTATCATCACTTACCAAGTACACAGTGGCAATACT	67633
QY	653	ARGASPMARGPHESERIILETHYRSEGINSEYGIYGIYLEUVALSERILEUALLEUAPRO	672
Db	67636	CGAAACCGCTTCTTCCATCAACAGCCAAAGGTGGTGGCTGTGTATCCCTTGGCCCGCA	67577
QY	673	LEUASPMYRILYSLEUGIUALARGGINUYRVALLEUALAVALTHYRALASERASPMGIYTHYRAG	692
Db	67576	CTGAGACTACAACTTGGCGGCGAGATGTGTGTGGCTGTATACCGCTTCCATGGCACTCGG	67517
QY	693	GLINASPMTHRALAGIINILEVALVALASPMVALTHYRASPALASPMTHRIISARGPROVALPHE	712

```

Db 67516 CAGGACACGCGACGATGTTGGATATGTCACCGACCGCAACCCATGCTGCTT 67457
QY 713 GlnSerSerHisGlyThrValAsnValAsnGluAspArgProAlaGlyThrThrValVal 732
Db 67456 CAGAGCTCCCACTATACATGTAATGTAATAGAGACCGGCGCGACGCGACCGAGG 67397
QY 733 LeuIleSerAlaThrAspGluAspThrGlyIleAsnAlaArgIleThrTyPheMetGlu 752
Db 67396 CTGATCAGCGCCACGATGAGGACACAGGTGAAGTCCCGCATCCACTTCATGAG 67337
QY 753 AspSerIleProGluPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlu 772
Db 67336 GACAGCATCCCGCAGTTCGCGATCGATGACACACGCGGCGCTGTCAACCCAGCGGTGAG 67277
QY 773 LeuAspTyGluAspGlnValSerTyThrLeuAlaIleThrAlaArgAspAsnGlyIle 792
Db 67276 CTGGACTACGAGACCAAGTCTTACACCTTGCCATTCATGCTCGGACATGGCAT 67217
QY 793 ProGlnTySerAspThrThrTyLeuGluIleLeuValAsnAspValAsnAspAsnAla 812
Db 67216 CCCGAGAAGTCCGACACCACTACCTGAGATCTGTGTAACGATGATGACATGCG 67157
QY 813 ProGlnPheLeuAspSerTyThrGlnGlySerValTyGluAspValProProPheThr 832
Db 67156 CCTCAGTTCCTCGAGACTCTTACCGAGGCGAGTGTATGAGAGATGTGCGACCTTCACT 67097
QY 833 SerValIleGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTy 852
Db 67096 AGCGTCTCCAGATCTCAGCCACTGATCGTGAATCTTGACCTTAAGGCGGGCTTCTAC 67037
QY 853 ThrPheGlnGlyIleAspAspGlyAspGlyIleValGlnSerThrSerGlyIle 872
Db 67036 ACCTTCGAAGGAGGAGCGATGAGACGGGTGACTTATGTTGAGTCCACGTCAGGCAATC 66977
QY 873 ValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyValLeuArgAlaTy 892
Db 66976 GGGCAACGCTACGAGGCTGATGAGAGAACGTCGACGATATGTCCTTGGCGCATAT 66917
QY 893 AlaValAspIleValIleMetProProAlaArgThrPrometGluValThrValIleu 912
Db 66916 GCAAGTGCACAAAGGAGATGCCGCCACGCCGACACCTTAAGAGTACAGTCTGTGTTG 66857
QY 913 AspValAsnAspAsnProProValPheGlnIleAspGluPheAspValPheValGlnGlu 932
Db 66856 GATGAGATGACAAATCCCTGCTTGAAGCAGATGATGATGATGTTGTGTGGAAGG 66797
QY 933 AsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThr 952
Db 66796 AACACCCCTATGGGTACGCGTGGCCCGGATCACAGCCACGACCCGATGAGGCAAC 66737
QY 953 AsnAlaGlnIleMetTyGlnIleValIleGluIleValIleProGluValPheGlnIleuAsp 972
Db 66736 AATGCCAGATATATGATCAGATTTGTGAGGGGCAACATCCCTGAGGTCTTCAAGCTGAG 66677
QY 973 IlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyGluAspArgProGluTy 992
Db 66676 ATCTTCTCCGGGAGACTGACACCTCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 66617
QY 993 ValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArg 1012
Db 66616 GTCCCTGATCAGGCGCACGTCAGCTCTCTGAGGAGCGGCGGCTACAGTCCACCTCGC 66557
QY 1013 LeuLeuAspArgAsnAspAsnProProValIleGluIleAspPheGluIleLeuPheAsn 1032
Db 66556 CTCTTGAACCGAAGACAAACCAACCAAGCTGGGCAATCTTTTCAACAAAC 66497
QY 1033 TyrValIleAsnArgSerSerPheProGlyIleGlyIleValArgValProAlaHis 1052
Db 66496 TATGTACCAATCGCTCAAGACGCTTCCCTGGGCGGTGCAATGGCGAGTACTCCCAT 66437
QY 1053 AspProAlaIleSerAspSerIleuThrTySerPheGluArgGlyIleGlnIleuSerIleu 1072

```

```

Db 66436 GACCTGATATCTCAGATATGCTGACTTACAGCTTTGAGCGGGAATGATCACTG 66377
QY 1073 ValLeuLeuAsnAlaSerThrGlyIleuTySerArgAlaIleuAspAsnArg 1092
Db 66376 GTCTGCTCAATGCTTCCACGGGTAGCTGAAGCTTACCGCGCATGCAACACCG 66317
QY 1093 ProLeuGlnAlaIleMetSerValLeuValSer 1103
Db 66316 CCTCTGAGGCGCATATGAGCGTGTGCTGCTCA 66284

RESULT 12
AK091437 2821 bp mRNA linear PRI 15-JUL-2002
LOCUS AK091437
DEFINITION Homo sapiens cDNA FLJ34118 f18, clone FCBBF3009428, highly similar
to Homo sapiens FLAMINGO 1 mRNA.
ACCESSION AK091437.1 GI:21749808
VERSION AK091437.1
KEYWORDS oligo capping, f18 (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T.,
Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R.,
Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, B., Momiyama, H.,
Onogawa, S., Kaeriyama, S., Satoh, N., Matsumawa, H., Takahashi, E.,
Karaoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S.,
Terashima, Y., Matanabe, M., Sugiyama, T., Irie, R., Otsuki, T.,
Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai, H., O., Y., Salto, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahara, K., Maehno, Y., Nagai, K. and Isegai, T.
NEBO human cDNA sequencing project
Unpublished
2 (bases 1 to 2821)
Isegai, T. and Yamamoto, J.
REFERENCE
AUTHORS Isegai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isegai, FJI Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEBO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..2821
location/qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FCBBF3009428"
/tissue_type="brain"
/clone_lib="FCBBF3"
/dev_stage="fetal"
/note="cloning vector: pMB18SFL3"
BASE COUNT 540 a 877 c 835 g 569 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2821
Score: 609.00 Matches: 782
Percent Similarity: 99.62% Conservative: 0
Best Local Similarity: 99.62% Mismatches: 2
Query Match: 20.83% Indels: 3
DB: 9 Gaps: 0
US-09-916-849a-3 (1-2923) x AK091437 (1-2821)

```

QY 1116 ArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuAsp 1135  
 DB 231 CGATGACCATCATCAACGATAGATGCTCAACCAAGCATCAAGCGCCCTGGAGAC 230  
 QY 1136 MetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAla 1155  
 DB 291 ATGTGACCCGAGCGCTTCTGTCAACACTGTAGCCCTTTCATCCAGCGGTGGCGCC 350  
 QY 1156 ThrLeuAlaThrProAspPheValValPheAsnValGlnArgAspThrAspAla 1175  
 DB 351 ACGTGGCCACGCAACCGGACCACTGTGTCTTCAACCTACAGCGGACACCGAGCC 410  
 QY 1176 ProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyProGly 1195  
 DB 411 CCGCGGGCCCAATCCTCAACGTGAGCTGTGGTGGCCAGCGCCAGGGCCGGGGGC 470  
 QY 1196 GlyProPheLeuProSerGlyAspLeuGlnGlyLeuGlyLeuValAsnArgSerLeu 1215  
 DB 471 GGGCGCCCTTCTGCTGTGAGGACCTGAGGAGCGCTTATACCTCAACCGAGCTG 530  
 QY 1216 LeuThrAlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGlu 1235  
 DB 531 CTGACGGCCATCTCGGACAGCGCTGTCTGCTTCAACGACATCTGCTGGGAG 590  
 QY 1236 ProCysGluAsnArgMetArgCysValSerValLeuArgPheAspSerSerAlaProPhe 1255  
 DB 591 CCGTGCAGAACTCATGCGCTGTGGTGTGGTGTGCTGCACTCTCCGCGCCCTTC 650  
 QY 1256 IleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArg 1275  
 DB 651 ATGCGCTCTCTCTCGTGTCTTCTCGGCCCATCAACCCGAGGAGCGCTGCGCGC 710  
 QY 1276 CysArgProGlyPheThrGlyAspArgCysGlyIleThrGlyValAspLeuCysArgSerArg 1295  
 DB 711 TGCCCGCCCGCTTCAAGGGTGACTACTGCGAGCCAGAGTGAACCTCTGCTACTCGCG 770  
 QY 1296 ProCysGlyProHisGlyArgCysArgSerArgGlyGlyGlyArgCysArgCysArg 1315  
 DB 771 CCGCTGTGCCCCCAACGGGCGCTGCGCGAGCGCGAGGGCGGCTTCACTCTGTGCT 830  
 QY 1316 AspGlyIleThrArgIleHisCysGlyValSerAlaArgSerArgIleAspArgCysArg 1335  
 DB 831 GATGGCTACACGGGTGAGCACTGTGAGGTGAGTCTGCTCAAGCCCTTGACAC-CCGGT 889  
 QY 1336 ValCysIleAsnGlyGlyIleThrCysValAsnLeuLeuValGlyGlyPheCysArgCys 1355  
 DB 890 GTCTGCAAGATGGGGCACTGTGTCACTGTGTGGCGGTTCAAGTGGATTC 949  
 QY 1356 ProSerGlyAspPheGlyIleProArgCysGlnValThrThrArgSerPheProAlaHis 1375  
 DB 950 CCAATCTGAGACTTCGAGAGCCCTACTGCGAGGTGACCAAGCCCACTTCCCGCCAC 1009  
 QY 1376 SerPheIleThrPheArgGlyIleuArgGlnArgPheHisPheThrLeuAlaLeuSerPhe 1395  
 DB 1010 TCTCTTACACCTTGTGGGCTGTGCGGACAGGTTTCACTTCAACCTTGGCCCTTCTCTT 1069  
 QY 1396 AlaThrArgGluArgAspGlyLeuLeuLeuIleThrAsnGlyArgPheAsnGlyValAsp 1415  
 DB 1070 GCCACAAAGGAGCGCGAGCGGTGTGTGTGACATGGCGTTTCAATGAGAGCAGAC 1129  
 QY 1416 PheValAlaLeuGluValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyIleSer 1435  
 DB 1130 TTTGTGACCTCTCGAGGTGATCCAGAGACAGGTCCAGCTCACTTCTCTGAGGGGAGTCA 1189  
 QY 1436 ThrThrThrValSerProPheValProGlyGlyValSerAspGlyIleThrPheIleVal 1455  
 DB 1190 ACCACACGCGTGTCCCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1249  
 QY 1456 GlnLeuArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 1475  
 DB 1250 CAGGTAAATACCTCAATTAAGCCACTTGTGGTCAAGAGGCTCCACAGGGCCCATCA 1309

QY 1476 GlnGlnCysValAlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPhe 1495  
 DB 1310 GAGCAGAGAGT 1369  
 QY 1496 GlySerValLeuGlyValAsnArgSerCysAlaAlaGlnGlyIleThrGlnGlyGlySerIleVal 1515  
 DB 1370 GGAATGTCTGTGGCAATCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1429  
 QY 1516 SerLeuAspLeuThrGlyProLeuLeuLeuGlyValProAspLeuProGlySerPhe 1535  
 DB 1430 TCTCTGATCTGACGGGGCCCTGTCTACTAGCGGGGTCTGACTGCTGCCAGAGCTTC 1489  
 QY 1536 ProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHis 1555  
 DB 1490 CCAATCCGAATCGGAGT 1549  
 QY 1556 IleAspMetAlaAspPheIleAlaAsnArgIleThrValProGlyCysProAlaValVal 1575  
 DB 1550 ATTAGATGTGTACTTCACTTGTCCAAATGGGACCTGTCTGTGTGTGTGTGTGTGTGTGT 1609  
 QY 1576 AsnValCysAspSerAsnThrCysHisAsnGlyIleThrCysValAsnGlyIleThrAspAla 1595  
 DB 1610 AACGTGTGTGACAGCAACCTTGCACAAATGGGGGCACTTGTGTGACCAAGTGGAGCGG 1669  
 QY 1596 PheSerCysGlyCysProLeuGlyPheGlyGlyIleValSerCysAlaGlnGlyMetAlaAsn 1615  
 DB 1670 TTCACTGTGAGATGTGCCCCCTGTGGGCTTGTGGGCAAGAGCTGTGCCCAAGAAATGGCCAA 1729  
 QY 1616 ProGlnHisPheLeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSer 1635  
 DB 1730 CCAAGACACTCTCTGTGGGAGAGAGCTGTGGCTGTGCAATGGGCTCTGCTGCCCATCTCC 1789  
 QY 1636 GlnProThrPheLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGln 1655  
 DB 1790 CAACCTGTGTACTCAAGCTTCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1849  
 QY 1656 AlaIleThrArgIleArgSerThrIleThrLeuGlnLeuArgGlyHisValMetLeu 1675  
 DB 1850 GCCATCAACAGGGGGGAGAGCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1909  
 QY 1676 SerValGlnGlyIleThrGlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyIleArgAlaAsn 1695  
 DB 1910 AGGTGTGAGGAGCAAGGCTTCAAGGCTTCTCTCTCTGCTGTGAGCCAGGCGGCGCAT 1969  
 QY 1696 AspGlyAspThrHisIleAlaGlnLeuAlaLeuGlyAlaSerGlyIleProGlyHisAla 1715  
 DB 1970 GACGGTGACTGGACACATGACAGG-GETGTGCACTGGGAGCAAGCGGGGCGCCGCGCATGC 2028  
 QY 1715 AlaLeuSerPheAspArgGlyGlnGlnArgAlaGlyIleAsnLeuGlyProArgLeuHis 1735  
 DB 2029 CATCTGTCTCTTCAATTAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088  
 QY 1735 GlyLeuHisIleLeuSerAsnIleThrValGlyIleProGlyProAlaGlyIleValAla 1755  
 DB 2089 TGTGTGTGACCTGTGAGCAATTAACAGTGGCGGAAATCTTGAGCCAGCGGGGTGTGAC 2148  
 QY 1755 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlyIleValAla 1775  
 DB 2149 CCGTGTCTTCCGGGGCTGT 2208  
 QY 1775 AsnLeuAspProSerHisGlyIleLeuSerIleAsnValGlnGlnGlyCysSerLeuProAs 1795  
 DB 2209 CAGCTGTGATCCAGAGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2268  
 QY 1795 ProCysAspSerAsnProCysProAlaAsnSerArgCysSerAsnAspThrAspSerArg 1815  
 DB 2269 CCTTGTGTACTAAACCGGT 2328  
 QY 1815 rSerCysSerCysAspProGlyIleArgIleValAsnArgCysThrAsnValCysAspLeuAs 1835  
 DB 2329 TTCCTGCACTGTGTATCCAGGT 2388  
 QY 1835 nProCysGlyHisGlnSerValCysThrArgIleArgProSerAlaProHisGlyIleThrCys 1855

Db 2389 CCGGTGTGACACCACTGTGTGACCCGCAAGCCAGTCCCAAGGCTATACCTG 2448  
 QY 1855 GGLUCYSPROFASMTYrLeuGLYProTYrCYSGIuThrArgIleASPGLnProCYSPR 1875  
 Db 2449 CCGAGTGTCCCAAAATTACTTGGGCAATCTGTGAGACCAAGATTGACCAAGCTTGTCC 2508  
 QY 1875 GATGGLYrTPrTGTGlyVHISProThrCYSGIYProCYSPaAnCYSPaValSerIYSGIYph 1895  
 Db 2509 CCGTGTCTGTGTGGAGACATCCACATGTGCCATGCACTGATGTGATGAGCAAAAGCTT 2568  
 QY 1895 eASPProASPProCYSP 1899  
 Db 2569 TACCCAGACTGC 2581

RESULT 13  
 AL672200/c  
 LOCUS AL672200 89129 bp DNA linear ROD 01-JUL-2002  
 DEFINITION Mouse DNA sequence from clone RP23-37619 on chromosome 3, complete  
 sequence.  
 ACCESSION AL672200  
 VERSION AL672200.10 GI:21684704  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 89129)  
 GARNER, P.  
 DIRECT SUBMISSION  
 Submitted (01-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Jul 2, 2002 this sequence version replaced GI:21304370.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL, Sw,  
 SWISSPROT, Tr, TrEMBL, Wp, WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-37619 is  
 from the RCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

COMMENT  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL, Sw,  
 SWISSPROT, Tr, TrEMBL, Wp, WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-37619 is  
 from the RCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

## FEATURES

source

Location/Qualifiers

1. 89129

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="3"

/clone\_1ib="RP23-37619"

/clone\_1ib="RP23-37619"

BASE COUNT

23218 a 22283 c 22360 g 21268 t

ORIGIN

Alignment Scores:  
 Pred. No.: 0  
 Score: 482.00  
 Percent Similarity: 98.86%  
 Best Local Similarity: 98.86%  
 Query Match: 16.49%  
 DB: 10  
 Gaps: 0

US-09-916-849A-3 (1-2923) x AL672200 (1-89129)

QY 365 ValPheLeuSerValGluASPAsnAspAsnAlaProGlnPheSerGluYrArgTYR 404  
 Db 21678 GTTTCCTGTGTGGTGGAGATGATATATGACACGCCCCCATTTTATGACGAGAGGTTAT 21619  
 QY 405 ValValGlnValArgGluASPValThrProGlyAlaProValLeuArgValThraLaser 424  
 Db 21618 GTGTCTCAGTGTCCGAGAGATGTGACCCAGAGGCCCAAGTTCTCCAGTCAAGCCCTCC 21559  
 QY 425 AsparGAspLYGlySerAsnAlaVal-ValHisTYrSerIleMetSerGlyAsnAlaAr 444  
 Db 21558 GATAGAGCAAGGCGCAGCAATGCCCT-AGTGCATACAGCATCATGAGTGCATATGCTCG 21500  
 QY 444 GGLYGLnPhETyrLeuAspAlaGlnThrGlyAlaLeuAspValValSerProLeuAspTY 464  
 Db 21499 GGGCCAGTTCTATCTGTGATGCTCAGACCTGAGGCCCTGTGATGTGTAAGCCCACTGACTA 21440  
 QY 464 rGluThrThrLYSGIYrTYrThrLeuArgVal-ArgAlaGlnAspGlyGlyArgProProL 484  
 Db 21439 TGAAGCAACCAAGATATACCTTCGGAT-CCGGGCTCAGGACCGGTGGCTCTCCAC 21381  
 QY 484 eUSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProI 504  
 Db 21380 TCTCCAAATGCTCTGTGGCTAGTAAAGAGCAAGTCTTACATCAACATTAACGCCCCCA 21321  
 QY 504 LePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeuGlyTYrLeuV 524  
 Db 21320 TCTTGTGACACCCCTTTCACAGCGCACCTCTTGAAGAGTGTCTCTTATGCTACCTTG 21261  
 QY 524 alLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTYrArg- 543  
 Db 21260 TTCTCCAGTCCAGGCAATTGATGCTGACCTGTGGATTAACGCCCTCCAGATTAAG-C 21202  
 QY 544 LeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTYrGlyTYrIleSer 563  
 Db 21201 CTGGCAGGGGTGGGACGACATTTCTCCCTTCAACATTAACAATGACAGGCTGATCTCT 21142  
 QY 564 ValAlaAlaGluLeuAspArgGluGluValAspPheTYrSerPheGlyValGluAlaArg 583  
 Db 21141 GTGGCCGACAGATTGGATCGGAAGAGGTGATTTCTACAGCTTGTGTAGAGGCCCGG 21082  
 QY 584 AspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr-Val-LeuAspVa 603  
 Db 21081 GACCATGACACCCACGCGCTCACGCGCTCACGCACTGTGATGATACAT-CTTGAGCT 21023  
 QY 603 lAsnAspAsnAsnProThrPheThrGlnProGluTYrThrValArgLeuAsnGluAspAl 623  
 Db 21022 CATATACAAACCAACCACTTACAGCAGCAGAGTACACCGTTCGGCTCAACGAGAGCGC 20963  
 QY 623 alAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValI 643  
 Db 20962 GCGCCGTGGGACCAATGTGTGAGACCGTGTACGCGGTGATGACAGCGCTCAACAGTGTCT 20903  
 QY 643 eThrTYrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerG 663  
 Db 20902 TACCTACCAAGATCACCAAGCGGACACCGGACACCGCTTCTTATCAACAGCAAAAGTGG 20843  
 QY 663 YGlyGlyLeuValSerLeuAlaLeuProLeuAspTYrIYrLeuGluArgGlnTYrValle 683  
 Db 20842 TGGTGAAGTGTCTCTCTTGTGCTTACCGCTGACATCAAACTCCAGCGCGCATATGTCT 20783  
 QY 683 uAlaValThraLaserAspGlyThrArgGlnAspThraLaglnIleValAsnValTh 703  
 Db 20782 GGCACTGACTGCTCAATGACCAAGGACGACACACTCTCAATGATGATGATCTCAC 20723



QY 703 rAapAlaantHrHlaArgProValPheGInserSerHlaTyThrValAsnValaAngl 723  
 Db 20722 CGATGCGAACAACCCAGCTCCGCTTCCAGAGCTCCCACTACAGTCAAGCTATATGA 20663  
 QY 723 uAapAqPProAlaGlyThrThrValValleuIleSerAlaThrAspGluAphThrglYgl 743  
 Db 20662 AGACCGGCGACGAGGACCAACAGTGTGTGATCAAGCGCACAGATGAGAGACCGGGGA 20603  
 QY 743 uAaAlaAargIleThrTyRPhemGluApsSerIleProGlnPheArgIleAspAlaAs 763  
 Db 20602 GAATGCCCAATCACTTATATGAGAGATAGCATCCCACTTCCGAATCAAGCGGA 20543  
 QY 763 pThrgIyAlaValThrThrGlnAlaGluLeuApyrGlyuApsGlnValSerTyThrIle 783  
 Db 20542 CACTGGGGGCTGTCAACCAAGCTGAGTGAAGTCAAGAGACCAAGGTGCTTATACCTT 20483  
 QY 783 uAlaIleThrAlaAArgAspAsnGlyIleProGlnIlySerAspThrThrTyLeuGluI 803  
 Db 20482 GGCATCACTGTGCGGACATGGCATTCGCCAAGATGACACCACTATTGAGAT 20423  
 QY 803 eLeuValaAsnApyValaAsnApsAlaProGlnPheLeuAArgAspSerTyGlnIlySe 823  
 Db 20422 CCGTGGATGATGATGACATGACACCAAGCTCCCACTTCCAGAGATTCCTACAGGGCAG 20363  
 QY 823 rValTyrgIuApyValProProPheThrSerValLeuGlnIleSerAlaThrAspArgAs 843  
 Db 20362 CGCTACGAGAGAGTCCGCGCTTCAACAGCGTCTGACAGATCTCAGCCACGATGAGAA 20303  
 QY 843 pSerGlyLeuAangIyAArgValAphTyThrThrPheGlnGlyValApsArgIyApsGlyAs 863  
 Db 20302 CTCGGGTCTGAAGAGGAGGGGTCTTTCACACTTCCAAAGAGAGATGAGGGAGCGTGA 20243  
 QY 863 pPheIleValGluSerThrSerGlyIleValAArgThrLeuAArgLeuApsArgGluAs 883  
 Db 20242 CTTTATGTATGATCGACATCGGCAATGTGGCCACATCGGAGGCTGATGTATGAGAA 20183  
 QY 883 nValAlaGlnTyrValleuAArgAlaTyAlaValApyGlyMetProProAlaArgTh 903  
 Db 20182 CGTGGCCGAGTACGTCTGCGGCGCTATGCGGAGCAAGGGCATGCCACACCGCGCAC 20123  
 QY 903 rPProMetGluValThrValThrValleuApyValaAsnApsProProValPheGluI 923  
 Db 20122 ACCCAAGAGAGTACAGTATCTGTCTCGAGTGTGAGTGAAGACACCCCTGTCTTGAACA 20063  
 QY 923 nApsGluPheApyValaPheValGluGluAAsnSerProIleGlyLeuAlaValaArgVa 943  
 Db 20062 GATGAGCTTGAAGTATTTCTGTAAGAGAACACCCCATCGGCTGGCCCTGCGCGGT 20003  
 QY 943 lThrAlaThrApsProApsGluGlyThrAsnAlaGlnIleMetTyrgIlnIleValGluG 963  
 Db 20002 CACAGCCACTGACCCAGATGAGAGACCAACGACAGATCATGTATACAGATTTGTGAGGG 19943  
 QY 963 yAaNilleProGluValPheGlnIleuApyIlePheSerGlyGluLeuThrAlaIleuVala 983  
 Db 19942 CAATATCCCTGAGAGTCTTCAAGTGAATATATCTCTGCGAGCTAACGCTCGGTGA 19883  
 QY 983 pLeuApyTyrgIuApsArgProGluTyrValleuValIleGlnAlaThrSerAlaProle 1003  
 Db 19882 TTTGAGACTAGAGAGCCGCGCTGAATATGTCTGTGATCAAGGCTACGCTCTCTCTT 19823  
 QY 1003 uValSerAArgAlaThrValHlSValAArgLeuLeuApsArgAsnApsAndProProValle 1023  
 Db 19822 GGTGAGAGAGGCTACGTCTCATGTCCGCTCCCTTGACCGAGATATACCCGCGGTGCT 19763  
 QY 1023 uGlyAaNPheGluIleuApyPheAaApyTyAlaThrAaNPheSerSerPheProgl 1043  
 Db 19762 GGGGCACTTTGAGATTTCTTTAAACAATATGTATCAACACCGCTCCAGCACTTCCCGG 19703  
 QY 1043 yGlyAlaIleGlyAArgValProAlaHlaApsProApsIleSerApsSerleuThrTySe 1063  
 Db 19702 TGGTGCCATAGGCGGTGTGCTGCGCACGACCCGATATCTCAGACAGCTGACGTACAG 19643

QY 1063 rPheGluApyrgIyAangIuLeuSerIeuValleuLeuAaAlaSerThrgIyGluLeu 1082  
 Db 19642 CTTGAGGAGAGAAATGAATCAAGCTGTGTCTGCTCAATGCTCCACGCGTGAAGCTG 19585  
 RESULT 14  
 AC093365/C  
 LOCUS  
 DEFINITION  
 AC093365  
 Mus musculus clone RP23-60016, WORKING DRAFT SEQUENCE, 7 unordered  
 pieces.  
 AC093365  
 AC093365.2 GI:24182377  
 HTG, HTGS, PHASER1, HTGS\_DRAFT, HTGS\_FULLTOP.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 234932)  
 Birren, B., Nusbaum, C. and Lander, B.  
 Mus musculus, clone RP23-60016  
 Unpublished  
 2 (bases 1 to 234932)  
 Birren, B., Linton, L., Nusbaum, C., Lander, B., Ali, A., Allen, N.,  
 Anderson, S., Bana, N., Baerlein, V., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gadyne, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Haeferd, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamet, A., Karatas, A., Kelle, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marulis, N., Matthews, C.,  
 McCarthy, M., McGowan, P., McKernan, K., McPherson, R., Meldrum, J.,  
 Menus, L., Miroya, T., Mianga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nobu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Riebeck, M., Riley, R., Rice, C., Rogov, P.,  
 Roman, J., Roselt, M., Roy, A., Santos, R., Schauer, S., Schnapack, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Struss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submissions  
 Submitted (20-AUG-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 234932)  
 Birren, B., Nusbaum, C., Lander, B., Ali, A., Allen, N., Anderson, S.,  
 Bana, N., Baerlein, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gadyne, S., Gord, S., Graham, L., Grand-Pierre, N., Haefer, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamet, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
 Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Miroya, T.,  
 Mianga, V., Mianga, V., Naylor, J., Nguyen, C., Nicol, R., Nobu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunhang, P., Pierre, N., Raymond, C., Retta, R.,  
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnapack, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K.,  
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submissions  
 Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 21, 2002 this sequence version replaced g1:15213910.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center

TITLE  
 JOURNAL  
 COMMENT



Percent Similarity:	98.86%	Conservative:	0
Best Local Similarity:	98.86%	Mismatch:	4

Query Match:	16.49%	Indels:	8
DB:	2	Gaps:	0
US-09-916-849A-3 (1-2923) x AC093365 (1-234932)			
QY 385	ValPheLeuSerValGluAspAspAspAspAspAlaProGlnPheSerGlyIleValArgTyr		404
Db 139909	GTTTTCCTCGTCGGTGAAGATATATATGCAACGCCGCCCATGTTAGCGAAGCGTTAT		139850
QY 405	ValValGlnValArgGluAspValThrProGlyAlaProValLeuArgValIThrAlaSer		424
Db 139849	GTGGTCCAGGTGCGGGAGGATGTGAACCCAGAGGCCAGTTCTCCGAGTCAACGCTCC		139796
QY 425	AspArgAspGlyValGlySerAspAlaVal-ValHisTyrSerIleMetSerGlyAspAlaIar		444
Db 139789	GATTAAGACAAGGGGACGACATGCCCT-AGTGCATCAACATCATGATGATGGGCAATGCTCG		139733
QY 444	GGIYglnPheTyrLeuAspAspAlaGlnThrGlyAlaLeuAspValValSerProleuAspTyr		464
Db 139730	GGCCAGATTCTATCTGGATGCTCAAGCTGGAGCCCTGGATGGTGAAGCCCATCTGACTA		139677
QY 464	IGluThrThrIleArgGlyTyrThrLeuArgVal-ArgAlaGlnAspGlyGlyArgProProL		484
Db 139670	TGAGCAACCAAGATATATAGCTTCGAT-CCGGGCTCAGAGCGGTGGCGCTCTCCAC		139612
QY 484	euserAspAspValSerGlyLeuValIThrValGlnValLeuAspIleLeuAspAspAlaProI		504
Db 139611	TCTCCAAATGCTCTGGCCCTAGTAAACGTGCAAGTCTTAGACATCAACGATTAACGCCCCCA		139552
QY 504	IepheValSerThrProPheGlnAlaThrValLeuGluSerValProleuGlyTyrLeuVal		524
Db 139551	TCTTTGTGACAGCCCTTTTCCAGGCCACCGTCTCAAGAGTGTCCCTTAGGCTACCTTG		139492
QY 524	AlLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeuGlyTyrArg-543		543
Db 139491	TTCTGCACGTCCAGGCAATATGATGTGAACGTGGATGAATACGCCCGCTCGATATATG-C		139433
QY 544	LeuAlaGlyValAlaGlyHisAspPheProPheThrIleAspAspGlyIThrGlyTyrIleSer		563
Db 139432	CTGGCAGAGGGTTGGGCAAGACTTTCCTTCACATTAACAAAGGCAACAGCTGGATCTCT		139373
QY 564	ValAlaAlaGlyLeuLeuAspArgGluGluValAspPheTyrSerPheGlyValGlnAlaArg		583
Db 139372	GTGGCCGAGAGTTGGATCGGAGAGAGGTTGATTTCTACGCTTTGGTGTGAAGGCCCGG		139313
QY 584	AspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValIThrVal-LeuAspVal		603
Db 139312	GACCATGGCAACCCAGCGGCTCAGCGGCTCAGGCAAGTGCAGTGTAAACAT-CTTGACCGT		139254
QY 603	IAspAspAspAspProThrPheThrGlnProGlyTyrIThrValAlaGlyLeuAspGluAspAl		623
Db 139253	CAATGACAACAACCAACCTTCACGACGACCAAGAGTACACCGTTGGGCTCAACAGAGGACCG		139194
QY 623	AlaValAlaGlyThrSerValValIThrValSerAlaValAspArgAspAlaHisSerValIle		643
Db 139193	CGCCGTGGGCAACAGTGTGGTGAACCGTGTACGCGGTGATGAGACCGCTCAACAGTCTCA		139134
QY 643	eThrTyrGlnIleThrSerGlyAspThrArgAspArgPheSerIleThrSerGlnSerGly		663
Db 139133	TACCTACCAAGATCAACAGCGGCAACCCGCAACCGCTTCTTATCAACAGCCAAAGTGG		139074
QY 663	YgIYgIleuValSerLeuAlaLeuProleuAspTyrTyrIleGluIleuArgGlnTyrValIle		683
Db 139073	TGGTGAAGTGGTCTCCCTTGCTTACCGCTTAGACTACAAACTCGAGCGGCACTATGTGCT		139014
QY 683	uAlaValIThrAlaSerAspGlyIThrArgGlnAspIThrAlaGlnIleValValAspValITh		703
Db 139013	GGCAGTAGCTGCTCAAGTGCACAAGGCAAGACACAGCTCAAGATAGTGAATGTCAAC		138954
QY 703	rAspAlaAspThrHisArgProValPheGlnSerSerHisTyrIThrValAspValAspGly		723
Db 138953	CGAATGCCAACCCACCGTCCCGCTTCCAGAGCTCCCATCAACAGTCAACAGTCAAGTTATATCA		138894

QY 723 uAepAgtProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGly 743  
 Db 138893 AACCGCGCAGGAGGAGCAGCAGTGTGTGTGATCAAGCGCAGAGGAGAGCAGCGGGGA 138834  
 QY 743 uAenAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArgIleAspAlaAs 763  
 Db 138833 GAATGCCCAATCACTTACTTTATGAGGATAGCATCCCGAGTTCCGATTCGATGACCGGA 138774  
 QY 763 pThrGlyAlaValThrThrGlnAlaGluLeuAspArgGluAspGlnValSerThrThrle 783  
 Db 138773 CACTGGGCGCTGCACCAACCGAGCTGAGCTACGAGAGACCAAGTGTCTTATACCTT 138714  
 QY 783 uAlaIleThrAlaArgAspAsnGlyIleProGlnIlySerAspThrThrThrleuGluI 803  
 Db 138713 GGCCATCACTGCTCGGAGCAGATGGCAATCCCGAGAGTGTGACACCACTATTGTGAGAT 138654  
 QY 803 eLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerThrGlnGlySe 823  
 Db 138653 CCTGGTGAATGACGTAAACGACACGCCGCCAGTTCTCTAGAGATTCTTACAGGGGAG 138594  
 QY 823 rValIlyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThrAspArgAs 843  
 Db 138593 CGTCTACGAGAGAGTCCCGCTTCAACAGCGTCTGAGATCTCAGCCACGAGTGAAGA 138534  
 QY 843 pSerGlyLeuAsnGlyArgValPheThrThrPheGlnGlyGlyAspAspGlyAspGlyAs 863  
 Db 138533 CTCCGGTCTGAATGAGCAGGGGTTTTCTACACCTTCCAGAGAGAGATGGGAGCGGTGA 138474  
 QY 863 pPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAspArgGluAs 883  
 Db 138473 CTTTATTGTAGATCAACATCGGGCATTTGTGGCAGACCTGCGGAGCTGATCTGTAGAA 138414  
 QY 883 nValAlaGlnIlyrValIleuArgAlaTyraAlaAlaPlyGlyMetProProAlaArgTh 903  
 Db 138413 CGAGGCCAGTACGTCTCGGGGCTATGCGGTGAGCAAGGAGATCCACAGCGCCGAC 138354  
 QY 903 rProMetGluValThrValThrValIleuAspValAsnAspAsnProProValPheGluI 923  
 Db 138353 ACCCATGAGAGTACAGTACTGTCTGTGATGTGAATGACAAACCCCTGTCTTTGAAACA 138294  
 QY 923 nAspGluPheAspValPheValGluIuAsnSerProIleGlyLeuAlaValAlaArgVa 943  
 Db 138293 GATGTAGTTGACGTATTTCTGTGAAGAGAACGCCCATCGGGCTGGCCGCGCGGGGT 138234  
 QY 943 lThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrglnIleValGluI 963  
 Db 138233 CACAGCCACGACCCAGATGAAGGAGCAACGACAGATCATGTACCAAGATTGTGAGAG 138174  
 QY 963 yAsnIleProGluValPheGlnIleuAspIlePheSerGlyGluLeuThrAlaLeuValAs 983  
 Db 138173 CAATATCCCTGAGAGTCTTTCAGTTGATATATCTCTGCGAGACTAACCTGCTGTAGA 138114  
 QY 983 pLeuAspTyrgluAspArgProGluIlyrValIleuValIleGlnAlaThrSerAlaProle 1003  
 Db 138113 TTTGGACTATGAGGACCGGCTGATATATGCTGTGTCAATCCAGGCTACGCTCTCTCTT 138054  
 QY 1003 uValSerArgAlaThrValHiValAlaArgLeuAspArgAsnAspAsnProProValle 1023  
 Db 138053 GGTGACACGAGGCTACGTCTCATGTCCGCTCTTGACCGCAATGATTAACCGCGCTGCT 137994  
 QY 1023 uGlyAsnPheGluIleLeuPheAsnAsnTyraValThrAsnArgSerSerSerPheProI 1043  
 Db 137993 GGGCAACTTGTGAATCTCTTAAACAATATGTACCAACCGCTCCAGCAGCTTCCCGG 137934  
 QY 1043 yGlyAlaIleGlyArgValProAlaHiAspProAspIleSerAspSerIleuThrTyse 1063  
 Db 137933 TGTGTCAATAGGCGGTGTGCTGCCACGACCCGATATCTCAGACAGGCTGACGTACAG 137874  
 QY 1063 rPheGluArgGlyAsnGluLeuSerIleuValIleuLeuAsnAlaSerThrGlyGluLeu 1082  
 Db 137873 CTTTCAGCGAGAAATGAACTCAGCTTGTCTGTCTCAATGCTTCCACCGGTGAGCTG 137816  
 RESULT 15

AX098223  
 LOCUS AX098223 3912 bp DNA linear PAT 30-MAR-2001  
 DEFINITION Sequence 135 from Patent WO0118542.  
 ACCESSION AX098223  
 VERSION AX098223.1 GI:13515345  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Lee, J., Thompson, P. and Lillie, J.  
 TITLE Identification, assessment, prevention, and therapy of ovarian  
 cancer  
 JOURNAL Patent: WO 0118542-A 135 15-MAR-2001;  
 Millennium Predictive Medicine, Inc. (US)  
 FEATURES  
 source location/Qualifiers  
 1..3912  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 763 a 1256 c 1032 g 860 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 3912  
 Score: 347.00 Matches: 376  
 Percent Similarity: 99.47% Conservative: 0  
 Best Local Similarity: 99.47% Mismatches: 1  
 Query Match: 11.67% Indels: 2  
 DB: Gaps: 0  
 US-09-916-849A-3 (1-2923) x AX098223 (1-3912)  
 QY 2186 ArgTyrgluAlaLeuArgGlyGluGlnProProAspLeuGluThrThrValIleLeuPro 2205  
 Db 11 CGTTCAGAGGCTTCGCGGAGAGCAGCCCGAGACTTGAGACAAAGTCTTGCTT 70  
 QY 2206 GluSerValPheArgGluThrProProValValArgProAlaGlyProGlyGluAlaGln 2225  
 Db 71 GAGTGTCTTCAGAGAGACGCCCGCGGTGAGGCCCGGAGCGCCCGAGAGGCCAG 130  
 QY 2226 GluProGluGluLeuAlaArgGlnArgArgHisProGluLeuSerGlyGluAla 2245  
 Db 131 GAGCCAGAGAGCTGCGACCGGAGACGAGCGCACCCGAGCTGAGCCAGGTAGGCT 190  
 QY 2246 ValAlaSerValIleIleIleTyraArgThrIleuAlaGlyLeuLeuProHisAsnTyzAspPro 2265  
 Db 191 GTGGCCAGCGTATCATATCTACCGCACCTTGGCCGGGCTACTGCTCATATATGACCT 250  
 QY 2266 AspIlyAspSerLeuArgValProIlyArgProIleIleAsnThrProValIleSerIle 2285  
 Db 251 GACAAAGCCAGCTTGAAGATCCCAACGCCCGATCATCAACACCCGTGTGAGCATC 310  
 QY 2286 SerValHiAspAspGluGluLeuLeuProArgAlaLeuAspIlyProValThrValGln 2305  
 Db 311 AGCGTCAATGATGAGAGAGCTTCTGCCCGGCGCTGAGCAAAACCGTCAAGGTGAG 370  
 QY 2306 PheArgLeuLeuGluThrGluGluArgThrIlyProIleCysValPheThrAsnHisSer 2325  
 Db 371 TTCGCCCTGCTGAGACAGAGAGGAGCAAGACCATGTGTCTTCTGAGAACATTC 430  
 QY 2326 IleLeuValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGluValValPheArgAsn 2345  
 Db 431 ATCCGTGTCAAGGAGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 490  
 QY 2346 GluSerHisValSerCysGluCysAsnHisMetThrSerPheAlaValIleuMetAspVal 2365  
 Db 491 GAGAGCCAGTCAAGCTGCGCAGTGCACCAATGACAGCTTGTGTGTGTGTGTGTGTGT 550  
 QY 2366 SerArgArgGluAsnGlyGluIleuProLeuIlyThrLeuThrThrValAlaLeuGly 2385  
 Db 551 TCTCGGCGGAGAAATGGGAGATCTCTGCACTGAAGACATGACATACGTGCTCTAGGT 610



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 05:37:06 ; Search time 1422 Seconds

(without alignments)  
5548.845 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923

Sequence: 1 MRSPTGVPLPPPPPLLL.....ACTVDEDSGSRFLFPNRLH 2923

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-DB=/cgn2.1/USPTO-SP001-US09916849/runat.11022004.143916.21346/app.query.fasta.1.3079  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MTN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MTNLEN=0 -MAXLEN=200000000  
-USER=US09916849 -CGN 1 1 0 -runat.11022004.143916.21346 -NCP=6 -ICPU=3  
-NO MAP -LARGEDQUERY -NEG SCORES=0 -WAIT -DSPLLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: N\_Geneseq.13jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	2923	100.0	8772	22	AA511678	Human Flamingo cDN
2	2923	100.0	10531	25	ABZ42868	Human GPCR CERS2
3	2894	99.0	11762	25	ABX34546	Human mdtl cDNA SE
4	2837	97.1	8871	22	AA511677	Human Flamingo cDN
5	1977	67.6	9401	22	ABA08648	Human FLAMINGO 1 h
6	1958	67.0	9321	24	ABK15177	Human REPT9 cDNA
7	1623	55.5	9121	21	AACT6401	Human OREF ORF1956
8	776	26.5	2332	22	ABA19447	Human immune/haema
9	776	26.5	2332	22	AAK83060	Human immune/haema
10	776	26.5	2332	22	AA531490	Human DNA for a no
11	776	26.5	2332	24	ABQ66814	Human polynucleoti
12	729	24.9	2391	25	ABT31943	Human breast cance
13	691	23.6	2077	22	ABA19448	Human nervous syst
14	691	23.6	2077	22	AAK83061	Human immune/haema
15	691	23.6	2077	22	AA531491	Human DNA for a no
16	691	23.6	2077	24	ABQ66815	Human polynucleoti
17	537	18.4	2603	19	AAV07219	Human calcitonin r
18	409	14.0	1734	22	ABA19445	Human nervous syst
19	347	11.9	3912	22	AAF98728	Human late stage o
20	281	9.6	2695	24	ABL90074	Human polynucleoti
21	281	9.6	4152	21	AAFI5924	Human prostate can
22	220	7.5	3620	25	ABZ6319	Human secretory po
23	217	7.4	652	22	ABA19446	Human nervous syst
24	217	7.4	652	22	AAK83062	Human immune/haema
25	217	7.4	652	22	AA531492	Human DNA for a no
26	177	6.1	551	21	AAA4496	Human polynucleoti
27	177	6.1	551	21	AAA4496	Human secreted exp
28	131	4.5	509	22	AAK3059	Human nervous syst
29	131	4.5	509	22	AAK3059	Human immune/haema
30	131	4.5	509	22	AA511489	Human DNA for a no
31	118	4.0	509	24	ABQ66813	Human polynucleoti
32	118	4.0	509	22	AAI99622	Human expressed po
33	118	4.0	509	22	AA531252	Human cDNA encodin
34	118	4.0	658	24	ABQ66576	Human polynucleoti
35	105	3.6	443	22	ABA13396	Human nervous syst
36	105	3.6	443	22	AA531400	Human cDNA encodin
37	105	3.6	443	24	ABQ66724	Human genomic DNA
38	102	3.5	721	22	ABA06845	Human genomic DNA
39	102	3.5	721	22	ABA06845	Human genomic DNA
40	102	3.5	721	22	AA529019	Genomic sequence #
41	102	3.5	721	22	AA529019	Genomic sequence #
42	102	3.5	721	22	AA531656	Genomic sequence #
43	102	3.5	721	22	AA531657	Genomic sequence #
44	102	3.5	721	24	ABV84182	Human polynucleoti
45	102	3.5	721	24	ABV84183	Human polynucleoti

#### ALIGNMENTS

RESULT 1  
AA511678 standard; cDNA; 8772 BP.

XX AA511678;

DT 24-OCT-2001 (first entry)

XX Human Flamingo cDNA splice variant.

KW Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
KW signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
KW anorexia; bulimia; asclima; Parkinson's disease; acute heart failure; HIV;  
KW obesity; hypotension; hypertension; urinary retention; angina pectoris;  
KW myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
KW benign prostatic hypertrophy; psychotic disorder; neurological disorder;  
KW manic depression; delirium; dementia; severe mental retardation; ss;

KM Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KW antifungal; antiviral; antiprotzoal; anti-HIV; anorectic; antianorectic;  
 KW antiparkinsonian; cardiatic; cerebroprotective; neuroprotective;  
 KW antidepressant; anticonvulsant; antisense therapy; gene therapy.

OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..8772  
 FT /tag= a  
 FT /product= "Human Flamingo protein #2"

XX MO200161003-A1.

XX 23-AUG-2001.

XX 19-FEB-2001; 2001MO-GB00680.

XX 19-FEB-2000; 2000GB-0004196.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Testa TT;

XX WPI: 2001-502792/55.

XX P-PSDB; AA007054.

XX An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -

PS Claim 5; Page 29-33; 66pp; English.

XX The sequence represents a cDNA splice variant which encodes a human  
 CC Flamingo polypeptide. Flamingo is a member of the G-protein coupled  
 CC receptor family, which is involved in signal transduction pathways. By  
 CC screening to identify compounds that stimulate or inhibit the function or  
 CC level of the protein, treatments can be developed for various diseases  
 CC and bacterial, fungal, protozoan and viral infections, including HIV,  
 CC cancer, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's  
 CC disease, acute heart failure, hypotension, hypertension, urinary  
 CC retention, angina pectoris, myocardial infarction, stroke, ulcers,  
 CC allergies and benign prostatic hypertrophy. Also treatable are psychotic  
 CC and neurological disorders such as anxiety, schizophrenia, manic  
 CC depression, delirium, dementia, severe mental retardation, Huntington's  
 CC disease and Gilles de la Tourette's syndrome.

XX Sequence 8772 BP; 1696 A; 2814 C; 2564 G; 1698 T; 0 other;

XX Alignment Scores:

Pred. No: 0 Length: 8772  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-916-849a-3 (1-2923) x AAS1678 (1-8772)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 Db 1 ATGGGAGGCCCGGACCGGCGTCCCTCCCAACGCCCGCGCGCTGCTGCTG 60  
 QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCySarSerLeu 40  
 Db 61 TTGCTGCTGCTGCTGCGCGCGCACTATTGGGAGAACCAAGTGGGGCCCTGCTGCTT 120  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAlaProMetGlyTyrPleuCySarSer 60  
 Db 121 GGGGCCAGGGGACGAGCTCTTGGGGGCGCTGCGCCCAATGGGCTGCTGCTGCAATCC 180  
 QY 61 SerAlaSerLeuLeuTyrPleuTyrThrSerArgCySarGAspAlaGlyThrGlnLeuThr 80

Db 181 TAGGCTGGAACCTTGCTGCTTACACCAAGCGCTGCAGGATGCGGCACTGACTACT 240  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCySarProGlySerGlnAlaHis 100  
 Db 241 GGCACCTGTGATCCCAACAGATGGCTGAGGGTGTGGTGTCCAGAAATCCAGAGCCCAT 300  
 QY 101 IleProLeuProProAlaProGlnGlyCySarProTyrSerCySarGluLeuGlyIleGly 120  
 Db 301 ATTCCTCCCTAACCAACAGCTCTGAAAGCTGCCCCCTGAGCTGTGCGCTCTGGGCAATGGA 360  
 QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnHisProCySarLeuValAla 140  
 Db 361 GGCACCTTTTCCCAACAGGCAAGCTTCACTGCCCCGAGAGACACCTGCTTAAAGGCT 420  
 QY 141 ProArgLeuArgCySerGlnSerCySarValLeuAlaGlnAlaProGlyValLeuArgAlaGly 160  
 Db 421 CCACGGCTCAGATGCAAGTCCCTGCAAGCTGGGACAGGCCCGCGGGCTCAGGGGAGGGA 480  
 QY 161 ArgSerProGlnGlnSerLeuGlyValArgArgValArgAsnValAsnThrAlaProGln 180  
 Db 481 AGGTACACCAAGAAAGTCCCTGGGTGGGTGGGAAAGAAATGTAATACAGCCCCCAAG 540  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200  
 Db 541 TTCAGCCCCCAAGCTTACAGGCAAGTCCCGGAGAAACAGCCAGCAGGACCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlnTyrThrMet 220  
 Db 601 GCATCCCTGAGGGCCATGCACCCGGACAGGGGTGAGGCGATGCACTGGAGTACACCATG 660  
 QY 221 AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 661 GATGCCCTTTGATAGCCGCTCCCAACCAAGTCTTCTCCAGACCACTGACTGCTGCA 720  
 QY 241 ValThrThrAlaGlnGlnLeuAspArgGlnTyrThrLysSerThrHisValPheArgValThr 260  
 Db 721 GTACACCAAGCCGAGAGAGTGTGATGTGAGACCAAGACCAACCAAGTCTTCAAGGCTCAG 780  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 781 GCGCAGACCAAGGAGATGCCCGGACCAAGTGCCTTGCTACACTCACTCTTGTGTACT 840  
 QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlnTyrLysGlnSerLeuArgGln 300  
 Db 841 GACACCAATGACCAATGACCTGTGTTCAGACAGAGATCAAGAGAGAGCTCAGGGAG 900  
 QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGAGGTTGGCTATAGGCTGCTCACTGTCAAGGCCAAGATGTATGCCCTGCC 960  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyIleSerProSerGlnValPhe 340  
 Db 961 AATGCCAATATCTGTACCGCTGCTGAGGGGCTCGGGGGGACGCCCTCTGAAGCTTT 1020  
 QY 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnVal 360  
 Db 1021 GAGATCGAACCTCGCTGCTGGGGGTGATCCGAACCGGCGCTGTGATGAGGAAGGGTG 1080  
 QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1081 GATCTTACCACTGACGTGTAGAGCAAGTGCACAGGCTCGGGACCCGGGCTCTGGAGAT 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
 Db 1141 ACCACAGCCCGCTGTTTCTTCTGTGTGGAGATGCAATGAATAAATGCCCCCAAGTTTACT 1200  
 QY 401 GlnLysArgTyrValValGlnValArgGlnAspValThrProGlyValaProValLeuArg 420  
 Db 1201 GAGAGCGCTATGTGTGTCCAGGTAGAGAGAGATGTGATCTCCAGGGGCCCAATATCTCCA 1260  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSerIleMetSer 440  
 Db 1261 GTCAAGCTCTCGGATCGAGACCAAGGGAGCAATGCGCTGTGTCACTATGCAATCATGAGT 1320

QY	441	GLYAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaLeuAspValAlaIleSer	460
Db	1321	GGCAAGCTCGGGGAGACAGTTTATCTGGATGCGCCAGACTGGAGCTCGATGATGGTGAAC	1380
QY	461	ProLeuAspDyrgLunThrLysGlyuTyrThrLeuArgValArgAlaGlnAspGlyGly	480
Db	1381	CCCTTGACCTATGAGACGACCAAGAGTACACCTTACGGGTGGACACAGATGGTGGC	1440
QY	481	ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp	500
Db	1441	CGTCCCCCACTCTCTATGCTCTGGCTTGGTGAAGTACAGTCCCTGATATCAACGAC	1500
QY	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu	520
Db	1501	AATGCCCATCTTCCCTGACAGACCCCTTCCAGGCTACTGTCCTGGAGAGCGCCCTTA	1560
QY	521	GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	1561	GGCTACCTGGTCTCCATGTCAGGCTATCGACGCTGATGCTGGTGAACAATGCCCGCTG	1620
QY	541	GlyTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560
Db	1621	GAATACCGCTTGGCTGGGGTGGGACATGACTTCCCTTCAACATCAATGACACAGGC	1680
QY	561	TrpIleSerValAlaAlaGlyLeuAspArgGlyGlyValAspPheTyrSerPheGlyVal	580
Db	1681	TGGATCTCTGGCTGCTGCTGAACCTGACCGGGAGAGATTGATTCTTACAGCTTTGGGGTA	1740
QY	581	GlyAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
Db	1741	GAAGCTCGAGACCATGGCACTCCAGCACTCACTGGCTCGCCAGTGTCAAGCTGATGTGC	1800
QY	601	LeuAspValAsnAspAsnAspProThrPheThrGlnProGlyTyrThrValArgLeuAsn	620
Db	1801	CTGGATGTCAAGACACACATCCACTTATCCCAACAGAGTACACATGGCGCTCAAT	1860
QY	621	GlyAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640
Db	1861	GAGAGTGCAGCTGTGGGACACAGCTGTGACCGTGTCACTGTGAGCCGTATGCTCAT	1920
QY	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer	660
Db	1921	AGTGTATATACCTTACAGATACACAGTGGCACTCGAAGACCGCTTCTTCATCACAGC	1980
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln	680
Db	1981	CAAGTGGTGGTGGCTGGTATCCCTTGGCCCTGGCACTGGACTGAATTAAGCGGCGAG	2040
QY	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700
Db	2041	TATGTGTTGGCTGTATACCCCTCCGATGCGACTCGGCAAGACACGCGCACAGATTGGTGTG	2100
QY	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720
Db	2101	AATGTACCGAGCGCAACACCACTGCTGTCTTCAAGCTCCCACTATACAGGAT	2160
QY	721	ValAsnGlyAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlyAsp	740
Db	2161	GTTTAATGAGGACCGGCGGCGGACACACGCTGGTCTGATCAAGCCGACGATAGAGAC	2220
QY	741	ThrGlyGlyAsnAlaArgIleThrTyrPheMetGlyAspSerIleProGlnPheArgIle	760
Db	2221	ACAGGGAGAAATGCCCGCATCACTTACTATGAGAGACAGCATCCCAAGTTCCGCATC	2280
QY	761	AspAlaAspThrGlyValAlaValThrThrGlnAlaGlyLeuAspTyrGlyAspGlnValSer	780
Db	2281	GATGCAAGACCGGGGCTGTCAACACCGCTGAGCTGAGCTGATCGAAGACCAAGTGT	2340
QY	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr	800
Db	2341	TACACCTCGGCATTTACTGCTGGGAGACATGGCATTTCCCAAGATGCCAGACACCTTAC	2400
QY	801	LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820
Db	2401	CTGAGATCTCTGGTGAACGACGTGAATGACAAATGCCCTCAGTTCCTGGAGACTCTTAC	2460
QY	821	GlnGlySerValTyrGlyAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	2461	CAGGGCAGTGTCTATGAGAGATGTGCACCTTCACTACAGCTTCAGATTCACACACT	2520
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly	860
Db	2521	GATCTGATTCCTGACCTTAATGACAGGCTCTTCACTCACTTCCAGAGGCGACATGGA	2580
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880
Db	2581	GACGGTGACTTTATGTTGAGTCCAGTCAAGCATCGTGGCAACCTTACGAGGCTGGAT	2640
QY	881	ArgGlyAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro	900
Db	2641	CGAGGAACGTGGCCCACTATGTCTTGGGCAATGACGTGGACAAAGGGGATGCCCA	2700
QY	901	AlaArgThrProMetGlyValThrValThrValLeuAspValAsnAspAsnProProVal	920
Db	2701	GGCCGCAACCTATGGAAGTGAAGTCACTGCTGTTGATGTGAATGACAAATCCCTGTTC	2760
QY	921	PheGlyAspArgLeuPheAspValPheValGlyGlnAsnSerProIleGlyLeuAlaVal	940
Db	2761	TTTGAAGCAGATGAGTTGATGTGTTGTGGAAAGAACAGCCCATTTGGGCTTACCGCTG	2820
QY	941	AlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyrGlnIle	960
Db	2821	GCCCGGTACAGCCACCTGACCCCGATGAGGACCAATGCCCAGATTATATGACAGATT	2880
QY	961	ValGlyGlyAsnIleProGlyValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla	980
Db	2881	GTCAGGGGCAACATCCCTGAGGTCTTCCAGCTGACACTTCTCCGGGAGCTGACACCC	2940
QY	981	LeuValAspLeuAspTyrGlyAspArgProGlyTyrValLeuValIleGlnAlaThrSer	1000
Db	2941	CTGGTAGACTTATGACTACAGAGACCGGCTGATGATCGTCTGATCATCAGGCCACGTCA	3000
QY	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuAspArgAsnAspPro	1020
Db	3001	GCTCTCTGGTGAAGCGGGCTACAGTCCAGCTCGCTTGAACCGCAATGACAAACCA	3060
QY	1021	ProValLeuGlyAsnPheGlyIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040
Db	3061	CCAAGTGGGCAACTTTGAAATCTTTTCAACAATATGTCAACAACTCGCTCAAGACGC	3120
QY	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	3121	TTCCCTGGGGGTGCCATTGGCCGAGTACCTGCCATGACCTGATATCTCAGATGTGCTG	3180
QY	1061	ThrTyrSerPheGlyArgGlyAsnGlyLeuSerLeuValLeuAsnAlaSerThrGly	1080
Db	3181	ACTTACACCTTGAAGCGGGAAATGAATCAAGCTCGTGTCTCAATAGCTTCCACGGGT	3240
QY	1081	GlyLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGlyAlaIleMetSerVal	1100
Db	3241	GAGCTGAAGCTTAAGCCGCGGACCTGAGACAAACCGGCTCTGGAGGCACTATGAGCCTG	3300
QY	1101	LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Db	3301	CTGGTGTACAGCGGGGTACACAGGCTGACCGCCAGATGCGGCTGGTCACTATCATTC	3360
QY	1121	ThrAspGlyMetLeuThrHisSerIleThrLeuArgLeuGlyAspMetSerProGlyArg	1140
Db	3361	ACCATGAGATGCTTCAACCAAGCATCAAGCTTGGCTTGGAGAGCATTCACCGAGGCGC	3420
QY	1141	PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Db	3421	TTCCGTGACACACATGTGATGCTCTTATTCAGGGGGGTGGCGGCAAGCTGGCCACGCA	3480
QY	1161	ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle	1180



D	3481	CCGACACACGTCGTCCTTCAACGTACAGGGGACCCGACGCCCCGGGGGACATC	3540	D	4561	GGGCCCCCTGCTACTAGCGGGGTGCTTGAACCTGCCCCGAGAGCTTCCAGTCCG	4620
Q	1181	LeuAnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu	1200	Q	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
D	3541	CTCAACGTAGACCTGTGGTGGGACAGCCGACAGGGCCCCGGGGGGGCGCCCTTCTTG	3600	D	4621	CAGTTCGTGGCTGACATCGGAACTTGCAGGTGAGACCCGGGACATTAACATGCTTGAC	4680
Q	1201	ProSerGlnAspLeuGlnGlyArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer	1220	Q	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIleValysAsnValCysAspSer	1580
D	3601	CCCTCTAGAGACCTGACGAGAGCGCTTACTCAACCGGACGCTGTCAGAGGCATCTCG	3660	D	4681	TTCAATTGCCAACATGACACCGGTGCTGCTCCCTGCCAAGAAAGACGTGTGACAGC	4740
Q	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGlyProCysGlyAsnTyr	1240	Q	1581	AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTTPAspAlaPheSerCysGlyCys	1600
D	3661	GCAACAGGCGGTGCTGCTTGCAGACAACTCTGCGCGGAGCCCTGGAGAACTAC	3720	D	4741	AACACTTGCACAAATGGGGCACTTGCATGAACCAAGTGGAGCCCTTCACTGCGAGTGC	4800
Q	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260	Q	1601	ProLeuGlyPheGlyGlyIlySerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu	1620
D	3721	ATGCGCTGCTGTGTGGTGTGCTGCTTGCATCTCCGCGCCCTTCAATGCTCTCTCC	3780	D	4801	CCCTTGCGCTTGGGGGCAAGAGCTGCGCCAGAAATGGCCATTCACAGCATTTCTCG	4860
Q	1261	ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe	1280	Q	1621	GlySerSerLeuValAlaTTPHisGlyLeuSerLeuProIleSerGlnProTTPTyLeu	1640
D	3781	GTGCTCTCCGGCCATCCACCCGTCGAGAGGCTGGCGCTGCGCCCGCCGCGCTTC	3840	D	4861	GGCAGACGCTGGTGGCTGGCATGGCTCTGCGTCCCATCTCCAACTTGGTACTTC	4920
Q	1281	ThrGlyAspTyrCysGlyThrGlnValAspLeuCysTyrSerArgProCysGlyProHis	1300	Q	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660
D	3841	ACGGGTACTACTGCGAGACCGAGGTGACCTCTGCTACTCGCGCCCTGTGGCCCCAC	3900	D	4921	AGCTCATGTTCGCAACGGCGCAGGCGCAAGGTGTCTGCTGCAAGCCATCCACAGGGG	4980
Q	1301	GlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly	1320	Q	1661	ArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGlyGlyThr	1680
D	3901	GGGCGCTGCGGACCCGAGAGGCGGCTACCTGCTCTGTGCTGTAATGGCTACACGGT	3960	D	4981	CGAGACACATCACCTTACAGCTTACAGAGGCGCACTGATCTAGCCCTGAGAGGACA	5040
Q	1321	GlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly	1340	Q	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyArgAlaAsnAspGlyAspTTPHis	1700
D	3961	GAGCACTGTAGAGAGTGCTCGCTCAGCGCGTTGCAACCCGAGGTGTCTGCAAGAATGG	4020	D	5041	GGGCTTCAAGCTCTCTCTCGCTGAGCGAGCGCGGCGCAATGACGTGACTGGCAC	5100
Q	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheGlyCysAspCysPProSerGlyAspPhe	1360	Q	1701	HisAlaGlnLeuAlaLeuGlyIleAsnGlyIleProGlyHisAlaIleLeuSerPheAsp	1720
D	4021	GGCACCTGTGTCAACCTGTGTGGGCGTTTCAATGACCAATGCCATCTGAGACTTC	4080	D	5101	CATGACACCTGTGACATGGAAGCAGGGGGGAGCTGGCGCAATGCTGTCTTCAT	5160
Q	1361	GlyIlyProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380	Q	1721	TyTGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisGlySer	1740
D	4081	GAGAGACCTTACTCCAGGTACCAACGCGAGCTTCCCGCCCACTCTTCAATCACTTT	4140	D	5161	TATGGCACAAGAGACAGAGGCAACCTGGGCCCGCCGCTGCATGTCTGACCTGAGAC	5220
Q	1381	ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIlyGlyArg	1400	Q	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyIleValAlaArgGlyPheArgGly	1760
D	4141	CGCGGCTGCGCGCAGCTTTCATCTTCACTGACCTGGCCCTCTGTGTCACAAAGAGACGC	4200	D	5221	AACATAACGTGGGGCGGAATACCTGGGCCAGCGGGGTGTGGCCCTTTCGCGGC	5280
Q	1401	AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIlyHisAspPheValAlaLeuGly	1420	Q	1761	CysLeuGlnGlyValArgValSerAspThrProGlyValAsnSerLeuAspProSer	1780
D	4201	GACGGTGTGCTGTGTCAATGGGCGTTTCAATGAGAAACATGACTTGTGGCCCTCGAG	4260	D	5281	TGTTTCAGGGGTGTGGGTGAGCCATACGCCAGAGGGGTTTAAACGCTTGATCCAC	5340
Q	1421	ValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrThrValSer	1440	Q	1781	HisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
D	4261	GTCATTCAGAGAGAGCTCAGCTCACTCTCTGACAGGAGAGTACACACACGGTCTCC	4320	D	5341	CATGGGAGAGACATCAACGTGGAGCAAGGCTGTAGCTGCTGACCTTGTGACTCAAAC	5400
Q	1441	ProPheValProGlyGlyValSerAspGlyGlyIleTTPHisThrValGlnLeuTyrTyr	1460	Q	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTTPAspSerTyrSerCysSerCysAsp	1820
D	4321	CCATTCTGCGCCGAGAGAGTCACTGATGGCCATGAGCATAGCGGTGACGTGAATACTAC	4380	D	5401	CCGTGTCTGTAAACGCTATTGCAAGCAACGCTGGGACAGCTATTCCTGCGACTGTGAT	5460
Q	1461	AsnIlySerProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnIlyValAla	1480	Q	1821	ProGlyTyrThrGlyValAspAsnCysThrAsnValCysAspLeuAsnProCysGlyHisGln	1840
D	4381	AATTAAGCAGCTGTGGGTGACAGAGGCTCCACAGGGGCCATCAGAGCAGAAAGGTGGCT	4440	D	5461	CAGGTACTAGGTGTACAACTGTACTAATGTGTGAACTTGAAACCGGTGTAGACACAG	5520
Q	1481	ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly	1500	Q	1841	SerValCysThrArgIlyAspSerAlaProHisGlyTyrThrCysGlyCysProProAsn	1860
D	4441	GTGCTGACCGTGTGATGTGTGACACAGAGTGGCTTGGCGCTTGGATCTGTCTGGGC	4500	D	5521	TCTGTGTGATCCGGCAAGCCAGTGCCTCCCATAGGTATATCTGAGAGTGTCCCCAAT	5580
Q	1501	AsnTyrSerCysAlaIleGlnGlyThrGlnGlyIlySerIlyValysSerLeuAspLeuThr	1520	Q	1861	TyrLeuGlyProTyrCysGlyThrArgGlyIleAspGlnProCysProArgGlyIlyTTPGly	1880
D	4501	AACACTCTCTGTGTGCGCCAGAGGACCCAGGTGGAGCAAGAGTCTGTGAGACTTACG	4560	D	5581	TACCTTGGGCAATACGTGTGAGACAGAGATTGACACACCTTGTCCCGGTGGCTGGGGA	5640
Q	1521	GlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMetArg	1540	Q	1881	HisProThrCysGlyProCysAsnCysAspValSerIlyGlyPheAspProAspCysAsn	1900
D				D	5641	CATCCACATGTGGCCCATGCACTGTGATGTACGAAAGCTTTGACCAAGACTGCAC	5700

OY	1991	lysthrSerSerglyGluCysHisGlySerGluLeuHisGlyPheArgProGlyIleSerProThr	1920
Db	5701	AAAGAAACACGGGAGAGTGGCACTGCAGAGAGAAACCACTACCGGCCCCAGGAGACGCCCAAC	5760
OY	1921	CysLeuLeuCysAAspCysSerTyPProThrGlySerLeuSerAArgValCysAAspProGluAsp	1940
Db	5761	TGCCCTCTGTGTGATCGTCTGATACCCCAAGGCTCTTGTCTGACAGTGTGTGACCCCTGAGGAT	5820
OY	1941	GlyGlnCysProCysAlaSerProGlyValIleGlyArgGlnCysAAspArgCysAAspAsnPro	1960
Db	5821	GGCCAGTGTCCATGACCAAGCCAGGTGTCACTCGGCGCTCAAGTGTGACCGCTGTGACAACTCT	5880
OY	1961	PheAlaGluValAlaThrThrAsnGlyCysGluValAsnIleTyPAspSerCysProArgAlaIle	1980
Db	5881	TTGCTGTAGGTGATCCACCAATGGCTGTAAAGTAATTAAGACGCTGCCACAGCGCAATT	5940
OY	1981	GluAlaGlyIleIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysArgPro	2000
Db	5941	GAGGCTGGGAGTGTGGTGGCCCGGTACCCGGCTTCGGGCTGGCTGTCTGCTCTCTGTCTCC	6000
OY	2001	LyseGlySerPheGlyValThrAlaValArgHisCysAAspArgIleAsArgIleTyrTrpLeuProRo	2020
Db	6001	AAAGGCTCTTTTGGAGACTGCTGTGGCCACTGTGATGACACAGGGGGGTGGCTCCCCCA	6060
OY	2021	AsnLeuPheAsnCysThrSerSeriLeThrPheSerGluLeuArgGlyPheAlaGluArgLeu	2040
Db	6061	AACCTCTTCAACTGCACGCTCCATCACTTCTCAAGAACTGAAGGGGCTTGCTGAAGCGGCTA	6120
OY	2041	GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg	2060
Db	6121	CAGCGGAATGAGTACAGGCTGACCTCAAGGCGCTCCACAGACGTACGCTCTCTCTGGGC	6180
OY	2061	AsnAlaThrGlnHisThrAlaGlyTyPArgGlySerAspValIleValAlaIleTyrGlnLeu	2080
Db	6181	AACGCGACAGACACACAGCTGTGCTACTTCGGACGACGCTCAAGGTGGCTTACCAAGCTG	6240
OY	2081	AlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln	2100
Db	6241	GCCACGCGGCTGTGGGCCACAGAGAGACCCAGCGGGGCTTTGGGCTGTCTGCACACAG	6300
OY	2101	AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
Db	6301	GACGCGCACTTCACTGAGAACTGCTGTGGGGGAGCGCGCTCTGTGACACAGCCCAAC	6360
OY	2121	LyAsnArgHisThrProGluLeuIleGlnGlnThrGlnGlyGlyThrAlaAlaTrpLeuLeuGlnHis	2140
Db	6361	AAGCGGCACTGGAGGCTGATCAGAGACAGAGGGGTGGCACCGCTGCTGCTCAAGAC	6420
OY	2141	TyrGlnAlaIleTyrAlaSerAlaLeuAlaGlnAsnMetArgHisSerThrTyPLeuSerProPhe	2160
Db	6421	TAATAGGCTTACGCGCAGTGGCCCTTGCCACAGAACATGGCGCACCACTTCAAGCGCCCTTC	6480
OY	2161	ThrIleValAlaThrProAsnIleValIleSerValAlaArgLeuAspLyseGlyAsnPheAla	2180
Db	6481	AACATCGTCAGCGCCCAACTTGTATCTCCGTAAGTGGCTTGGACAAAGGAACTTTGCT	6540
OY	2181	GlyValAlaLyseLeuProArgTyrGlnAlaLeuArgGlyGlyGlnIleProProAspLeuGlnThr	2200
Db	6541	GGGGCGAAGCTGGCCCCGTACAGAGCCCTGTGTGGGAGCGAGCCGCCGACCTTGAACA	6600
OY	2201	ThrValIleLeuProGluSerValPheArgGlnThrProProValAlaArgProAlaGly	2220
Db	6601	ACAAGTCATTCGCTGAGTCTGTCTTACAGAGAACGCCCCCGTGTGAGGCCCGCAGGC	6660
OY	2221	ProGlyGluAlaGlnGlnProGluGlnLeuAlaArgArgGlnArgArgHisProGluLeu	2240
Db	6661	CCCGAGAGGGCCACAGAGCCAGAGAGGTGTGCACGGCAGCGCAACCGCACTGCCAGACTG	6720
OY	2241	SerGlnGlyGluAlaValAlaSerValIleIleTyPArgThrLeuAlaGlyLeuLeuPro	2260
Db	6721	AGCCAGGCTGAGGCTGTGGCCAGGCTATCACTTACCGCACCTTGGCGGGGCTACTGCTCT	6780

QY	2261	HisAsnTYrAspProAspValAspSerLeuAspValProIysAspProIleLeuAsnThr	2260
Db	6781	CArNACTArVGACCTCGAAGACGAGCTTGAAGTCCCAACGCCCGATCATAAACA	6840
QY	2281	ProValValSerIleSerValHisAspAspGluIleLeuLeuProArgAlaLeuAspLys	2300
Db	6841	CCCGGGGTGACATCAGGCTCATGATGATGAGAGGCTTCGGCCCGGCGCTTGACAA	6900
QY	2301	ProValThrValGlnPheArgLeuLeuGluThrGluGluIleArgThrLysProIleCysVal	2320
Db	6901	CCCGTCACGGGACGAGTTCCCGGCTCTCGAGACACAGAGGCGGACCAAGCCCATCTGATGC	6960
QY	2321	PheThrPheAsnHisSerIleLeuValSerGlyThrGluGlyIleThrSerAlaArgGlu	2340
Db	6961	TTCTGGAACCATTTAAATCTGGTATGATGGACAGAGTGGCTGGATCGGCACAGGCTTGGA	7020
QY	2341	ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla	2360
Db	7021	GTGCTCTTCCCGCAATGAGAGCCGTCAGCTGGCAGTGACCAATGACGAGCTTGCT	7080
QY	2361	ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThr	2380
Db	7081	GTGCTCATGACGTTTCTCGCGGGGAAATGGGAGATCTGCGCATCTGAAGACATGACA	7140
QY	2381	TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu	2400
Db	7141	TACGGGGTCAAGATGTACCTTGCGTGGCTTCGTGCTCACTTCTTCTTCTCACTCTC	7200
QY	2401	LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly	2420
Db	7201	TTGGGTATCCCTGACCCCTCCAAACCAACAGGGATCCGATCACTGACAGGCTGGCTGGGC	7260
QY	2421	LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr	2440
Db	7261	CTGGCTCAGCTGGTCTTCTCTCTGGGAATCAACAGGCTGACTCTCTTTTGCTTGACA	7320
QY	2441	ValIleAlaIleLeuLeuHisIspheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGlu	2460
Db	7321	GTCAATGGCAATCCGTGTCACCTTCTGTACCTGTCGACCTTTTCTGGGCTGTGCTGGAG	7380
QY	2461	AlaLeuHisIleLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Db	7381	GCTTGCACCTGTACCGGGGACATCACTGAAGTGGCATGTATCAACCGGCCCATAGCGC	7440
QY	2481	PheTyrTyrMetLeuGlyIleTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Db	7441	TTCTACTACATGCTGGGCTGGGGGCTGCTCGCTTATGACAGGGGTACGCGTGGGCTG	7500
QY	2501	AspProGluGluTyrTyrGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuIle	2520
Db	7501	GACCCCGAGGGCTACCGGGAAACCTGTACTTGTGCTGTCCATCTATGACAGGCTCATC	7560
QY	2521	TyrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu	2540
Db	7561	TGAGATTTGCTGGCCCGGCTGGGCTTTGGCGTCCATGATGATGCTTCCCTGTACACCTG	7620
QY	2541	AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluLysGlyProValSer	2560
Db	7621	GCGGCCCCGGGCTCTGCTGTGCGCCAGCGGACGAGGCTTTGAGAGAAAGGTCTGTCCTCG	7680
QY	2561	GlyLeuGlnProSerPheAlaValIleLeuLeuLeuSerLysAlaThrThrLeuAlaLeu	2580
Db	7681	GGCTTGACGCTTCTTCGCGGCTCTCTCTGTGTGAGCCGACGATGGCTGTGGGACATG	7740
QY	2581	LeuSerValAsnSerAspThrIleLeuPheHisTyrIleLeuPheAlaThrCysAsnCysIle	2600
Db	7741	CTCTCTGTCAACAGCGACACCTCTCTTCCATCTACTCTTTCTGACCTGECATTTGCATC	7800
QY	2601	GlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGluValArgValAlaLeu	2620
Db	7801	CAGGGCCCCCTTCAATCTTCTCTCTATGATGGTCTTACAGAGAGGATCCGGAAAGCATC	7860
QY	2621	LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu	2640

```

Db      ||| 7861 AAGCTTGCTGACGCGCAGACCCACCTGACCTGCTGACCAAGTCCACCTCG 7920
Qy      ||| 2641 ThrSerSerTyranCyProSerProTyAlaAspGlyArgLeuTygInProTyArgLy 2660
Db      ||| 7921 ACCTGCTCTTCACTACCTGCCCCACCTTACGAGATGGGGGCTGTACCAAGCTTACGGA 7980
Qy      ||| 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyValSerSerInProSerTyrlle 2680
Db      ||| 7981 GACTCGCGCGCTCTCTGCAACAGCAACAGTGGCTCGGGCAAGAGTCAAGCCCAAGTCAATC 8040
Qy      ||| 2681 ProPheLeuLeuArgGluGluSerAlaLeuAspProGlyGlnGlyProProGlyLeuGly 2700
Db      ||| 8041 CCTCTTGTGTGAGGAGGAGGCTCGCACTGAACCTGCGCAAGGGGCCCCCTGCTCGGG 8100
Qy      ||| 2701 AspProGlySerLeuPheLeuGluGluSerAlaLeuAspProGlyGlnGlyProProGlyLeuGly 2720
Db      ||| 8101 GATCCAGGAGGAGGCTGCTTCTGAGAGGTCAGACCAAGCAAGCAATCTTACACGAGACTCC 8160
Qy      ||| 2721 AspSerAspLeuSerLeuGluAspAspGlnSerGlySerTyAlaSerThrHisSerSer 2740
Db      ||| 8161 GACAGTGAAGCTGCTTACAGAGACGACCAAGAGTGGCTCTTATGCTTACCCACTCATCA 8220
Qy      ||| 2741 AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760
Db      ||| 8221 GACAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280
Qy      ||| 2761 TrpAspSerLeuLeuGlyProGlyAlaGluArgLeuProLeuHisSerThrProLyAsp 2780
Db      ||| 8281 TGGGATAGCTGCTGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8340
Qy      ||| 2781 GlyGlyProGlyProGlyAlaProTyProGlyAspPheGlyThrThrAlaValGlu 2800
Db      ||| 8341 GGGGGGCCAGAGGCGCTGCAAGGCGCCCTGGCCAGAGACTTTGGAGCAACAGAAAGAG 8400
Qy      ||| 2801 SerSerGlyAsnGlyAlaProGlyGluArgLeuArgGluAsnGlyValSerAlaLeuSerArg 2820
Db      ||| 8401 AGTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8460
Qy      ||| 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisGlyLeuLeuGly 2840
Db      ||| 8461 GAGGGGCTCCCTAGAGCCCTTCCAGAGCTCTTCCAGAGCTCAAGAGGAGAGAGAGAGAGAG 8520
Qy      ||| 2841 LysGlyCysLeuProThrHisSerGlyValSerSerLeuLeuArgLeuProLeuGluGln 2860
Db      ||| 8521 AAGAAATGTCTGCCACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8580
Qy      ||| 2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880
Db      ||| 8581 TGCACAGGGCTTCCCGGGGCTCTCCGCTAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 8640
Qy      ||| 2881 ArgProProArgGlnSerLeuGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2900
Db      ||| 8641 CGCCCAAGCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8700
Qy      ||| 2901 SerTleValAlaGlyThrValAlaGluAspSerSerGlySerGluPheLeuPheLeu 2920
Db      ||| 8701 AGCATCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8760
Qy      ||| 2921 PheLeuHis 2923
Db      ||| 8761 TTCCTGAT 8769

```

RESULT 2  
AB242868  
ID AB242868 standard; DNA; 10531 BP.

XX  
AC AB242868;  
XX  
DT 04-MAR-2003 (first entry)  
XX  
DE Human GPCR CELSR2 nucleotide SEQ ID NO:523.  
XX

KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
KW G protein-coupled receptor modulator; antibody; immune-related disease;  
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
KW immunological-related cell proliferative disease; autoimmune disease;  
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
KW ulcer; gene; ds.

OS Homo sapiens.  
PN WO200261087-A2.  
XX  
XX  
PD 08-AUG-2002.  
XX  
XX  
PF 19-DEC-2001; 2001WO-US50107.  
XX  
XX  
PR 19-DEC-2000; 2000US-257144P.  
XX  
XX  
PA (LIFE-) LIFEESPAN BIOSCIENCES INC.  
XX  
XX  
PI Burner GC, Roush CL, Brown JP;  
XX  
XX  
DR WPI; 2003-046718/04.  
XX  
XX  
DR P-PSDB; ABP82018.  
XX  
XX  
PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
PT (GPCR), useful for diagnosing and designing drugs for treating  
PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
PT cancer or autoimmune diseases -  
XX  
XX  
PS Disclosure; Fig 1; 523pp; English.

CC The present invention describes antigenic peptides (I) comprising:  
CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-74 amino  
CC acids. Also described: (1) an assay for the detection of a particular  
CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity  
CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
CC an antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related disease, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention.

XX  
XX  
SQ Sequence 10531 BP; 2038 A; 3353 C; 2998 G; 2142 T; 0 other;

#### Alignment Scores:

Pred. No.:	Length:	10531
Score:	2923.00	2923
Percent Similarity:	100.00%	
Best Local Similarity:	100.00%	
Query Match:	100.00%	
DB:	25	
Gaps:	0	

US-09-916-849A-3 (1-2923) X AB242868 (1-10531)



Db 2223 GTTAATGAGGACCGGCGGACGACCAACGCTGCTGATACGCGCAAGATGAGAC 2282  
 Qy 741 ThGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheAlaGle 760  
 Db 2283 ACAGGTGAGATCCCGCATCCTCATGTAGAGACAGCATCCCGCATGCCATC 2342  
 Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db 2343 GATGACAGACCGGGGGGTGTCAACCAAGCTGAGTGAATGACAGACCAAGTGTCT 2402  
 Qy 781 TyrThrLeuAlaIleThrAlaArgAspArgIleProGlnIleSerAspThrTyr 800  
 Db 2403 TACACCTGGCCATTCTCTCGGGAACAATGCAATCCCAAGAGTCCGACACCACTAC 2462  
 Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyr 820  
 Db 2463 CTGGAGATCTGTGGTGAACGAGTGAATGACAAATGCCCTCAAGTTCCTGCGAGACTCTAC 2522  
 Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2523 CAAGGACGCTGTATGAGATGTGCCACTTCACTAGCGTCTTGAGATCTCAGCCACT 2582  
 Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly 860  
 Db 2583 GATCGTGAATCTGACTTAATGGCAGGGTCTCTACACTTCCAAAGAGGCGACGATGGA 2642  
 Qy 861 AspGlyAspPheIleValIleGluSerThrSerGlyIleValArgThrLeuArgGlyLeuAsp 880  
 Db 2643 GACGCTGACTTTATGTGTGATGACGTCAGGCAATCGGCAACGCTGCGAAGGCTGAGAT 2702  
 Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIleGlyMetProPro 900  
 Db 2703 CGAGAGAACTGGCCCGCATGTGTCTTGGCGCATATGCAATGACAAAGGAGATGCCCA 2762  
 Qy 901 AlaArgThrProMetGluValIleThrValIleValLeuAspValAsnAspAsnProProVal 920  
 Db 2763 GCCCGCACCTATGAGAGTGAACGATCATGTGTGTGATGATGAAATCAATCCCCCTGTC 2822  
 Qy 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 Db 2823 TTGAGAGAGATGAGTTTGAATGTGTGTGTGAGAGAGAACGCCCATTTGGGCTGACCGTG 2882  
 Qy 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 2883 GCCCGGATCACACCACTGACCCCGATGAAGGCAACAATGCCCATTAATGATCCAGATT 2942  
 Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyIleLeuThrAla 980  
 Db 2943 GTGGAGGCAACATCCCTGAGGCTTTCAGCTGACATCTTTCGGGGAGTGAACGCC 3002  
 Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000  
 Db 3003 CTGGTAGACTTAAGCTACGAGAACCGGCTGAGTACTCTGTGATCCAGGCCACACTCA 3062  
 Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 3063 GCTTCCTGTGTGAGCCCGGCTACAGTCCAGTCCGCTCTGTGACCCCAATGACACCCA 3122  
 Qy 1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040  
 Db 3123 CAGGTGCTGGGCAACTTGAATCTTTTCAACAATGATGACCAATCCCTCAAGAGAC 3182  
 Qy 1041 PheProGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 3183 TTCCCTGGGGGTGCTATGGCCAGTACTGCGCCATGACCTGTATCTCAGATAGTCTG 3242  
 Qy 1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3243 ACTTACGCTTTGAGCGGGAAATGAATCAAGCTGTGTCTGTCAATGTCTTCCACGGGT 3302  
 Qy 1081 GluLeuIleSerSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100

Db 3303 GAGCTGAAGCTAAGCCGCGCACTGGAACAACCCGCTCTGGAGCCATCATGACGCTG 3362  
 Qy 1101 LeuValSerAspGlyValHisSerValThrAlaGlnIleValAlaLeuArgValThrIleIle 1120  
 Db 3363 CTGGTGTGAGAGGGGTACACAGGCTGACCGCCAGTGGCCCTCGTGTGACATCATC 3422  
 Qy 1121 ThrAspGluMetLeuThrHisSerIleThrLeuAlaGlyLeuGluAspMetSerProGluArg 1140  
 Db 3423 ACCGATGAGATGCTCACCAACAGCATCAACGCTGGCCCTGGAGAGATGTCAACCGAGGCC 3482  
 Qy 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
 Db 3483 TTCTGTGACCACTGTCTAGGCTCTTTCATCCAGCGGGTGGCCGACGCTGGCCACGCCA 3542  
 Qy 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyValHisIle 1180  
 Db 3543 CCGAGACAGTGTGTCTTCAACGTTACCGGGACACGAGGCCCCCGGGGGCCACATC 3602  
 Qy 1181 LeuAsnValSerLeuSerValGlyIleProProGlyProGlyIleProProPheLeu 1200  
 Db 3603 CTCAAGTGAAGCTGTCCGGTGGCCAGCCGCGAGGGGCCGGGGCCGCTTCTCTG 3662  
 Qy 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
 Db 3663 CCTCTGAGGACTTCAGAGCGCTTATACCTCAACCGAGCTGTGACGGCATCTTG 3722  
 Qy 1221 AlaGlnArgValLeuProPheAspAsnIleCysIleAspGluProCysGluAsnTyr 1240  
 Db 3723 GCAAGCGGTGTCTGCTTCAACGACGAGGACCGGACCGGAGCCCTGGAGAACTTC 3782  
 Qy 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
 Db 3783 ATGCGCTGGGTGCTGATGTGGCTTTCGACTCTCCGCGCTTATGCGCTCTCTCC 3842  
 Qy 1261 ValLeuPheAspProIleHisProValGlyIleAspArgCysArgCysProProGlyPhe 1280  
 Db 3843 GTGCTCTTCGCGCCCATCAACCCGTCGAGGGCTGGCTGGCCGCGCCGCGCTTC 3902  
 Qy 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3903 ACGGGTGAATCTGGAGAACGAGGTGACCTGTCTACTGCGGGCCGTGTGGCCCCAC 3962  
 Qy 1301 GlyArgCysArgSerArgGluGlyIleTyrThrCysIleCysArgAspGlyTyrThrGly 1320  
 Db 3963 GGGCGCTGCGGACGCGGAGGGCGGCTAACCTGCTGTGATGCTGACACGGGT 4022  
 Qy 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGly 1340  
 Db 4023 GAGCACTGTAGAGTATGCTGCTCAAGCCCTGTGACCCCGGGTGTCTGCAGAAATGGG 4082  
 Qy 1341 GlyThrCysValAsnLeuLeuValGlyIlePheIleCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACTGTGCAACTGTGTGGGCGGTTTCAAGTGGATGTGCCCATCTGAGACTTC 4142  
 Qy 1361 GluIleProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 Db 4143 GAGAAAGCTTACTGCGAGGTGACACGCGCACTTCCCGGCACTCTTCACTCACTT 4202  
 Qy 1381 ArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAlaThrArgGluArg 1400  
 Db 4203 CCGGCTGCGGCAAGCGTTTCACTTCACTGACCTGGCCCTGTCTTGGCCCAAGAGAGGCC 4262  
 Qy 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIleAspPheValAlaLeuGlu 1420  
 Db 4263 GACGGGTGCTGTGTAACAATGGGCGTTTCAATGAGAAAGCAATGATTTGGCCCTCGAG 4322  
 Qy 1421 ValIleGlnGluValGlnLeuThrPheSerAlaGlyIleGluSerThrThrValSer 1440  
 Db 4323 GTGATCAAGAGAGGTCAAGCTCACTTCTGACGGGAAATCAACACAGGGTGTCC 4382  
 Qy 1441 ProPheValProGlyValValSerAspGlyIleTyrPheIleThrValGlnLeuIleTyrTyr 1460  
 Db 4383 CCAATGTGCGCGGAGAGTCAGTGTGCGCATACGCTGACAGTGAATATCTAC 4442



QY	1461	AsnlySPProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnValAla	1480	QY	1821	ProGlyTyrTyrGlyAspAsnCySerThrAsnValCyAspLeuAsnProCyGlnHISGln	1840
DB	4443	AATAAGCACATGTTGGGTCAAGACAGGCTCCACAGGGCCCATCAGACAGAGAGGTGCT	4502	DB	5523	CCAGTTACTATGTTGACAACTGACTAATATGTGTGACCTGAACCCGTGTGACACAG	5582
QY	1481	ValValThrValAsnGlyCyAspSerThrGlyValAlaLeuArgPheGlySerValLeuGly	1500	QY	1841	SerValCySerThrArglyProSerAlaProHISGlyTyrThrCySerGlyCyProProAsn	1860
DB	4503	GTGGTACCCGTGAGATGGCTGTGACACAGAGATGGCTTGGCTTCGATCTGTCTGGGC	4562	DB	5583	TCTGTGTACCCGCAAGCCCAAGTGGCCCCCAATGACTATGAGTGTGCTCCCAAT	5642
QY	1501	AsnTyrSerCyAlaAlaGlnGlyThrGlnGlySerTyrLeuSerLeuAspLeuThr	1520	QY	1861	TyrLeuGlyProTyrCySerGlnThrArgIleAspGlnProCyProArgIlyTyrTrpGly	1880
DB	4563	AACTACTCTGTGCTGCCACAGGGCACCCAGGGTGGCACAGAACTCTTGATTTGACG	4622	DB	5643	TACCTTGGGCAATCTGTGAGACCAAGATTGACCAAGCTTGTCCCGTGGTGGGGA	5702
QY	1521	GlyProLeuLeuLeuGlyGlyValProAspLeuProGlnSerPheProValArgMetArg	1540	QY	1881	HISProThrCySerGlyProCyAsnCyAsnValSerLySGlyPheAspProAspCyAsn	1900
DB	4623	GGGCCCCCTGACTAGCGGGGTGCTGACCTGCCAGAGCTTCCAGTCCGAATGCGG	4682	DB	5703	CATCCCAATGTGGCCATGCACTGTGATGTGACAAAGCTTTGACCCAGATGCAAC	5762
QY	1541	GlnPheValGlyCyMetArgAsnLeuGlnValAspSerArgHISIleAspMetAlaAsp	1560	QY	1901	LySerSerGlyGlyCySerIleGlyGlnAsnHISGlyTrpArgProProGlySerProThr	1920
DB	4683	CAGTTCGGGCTGATCGAGAACTGACAGGTGACACCGGCAATGACATGATGCTGAC	4742	DB	5763	AAGCAAGCGGCACTGCCACTGCAAGAGAACACTACCGGCCCAAGGACGCCAC	5822
QY	1561	PheIleAlaAsnAsnGlyThrValProGlyCyProAlaLySlyAsnValCyAspSer	1580	QY	1921	CysLeuLeuCyAspCySerTyrProThrGlySerLeuSerArgValCyAspProGlnAsp	1940
DB	4743	TTCAATGCCAACATAGGACCGGTGCTGGCTGCTCCCTGCCAAGAACATGTGTGACAGC	4802	DB	5823	TGCTCTTGTGTGACTGTACCCCAAGGCTCTTGTCCAGAGTCTGTGACCTGAGAT	5882
QY	1581	AsnThrCySerHISAsnGlyGlyThrCyValAsnGlnTyrAspAlaPheSerCySerGlyCyS	1600	QY	1941	GlyGlnCySPProCySlySPProGlyValIleGlyArgGlnCyAspArgCyAspAsnPro	1960
DB	4803	AACTTGCACAAATGGGGGACCTTGCGTAACAGTGGAGACGCGTTGACGTGCGAGTGC	4862	DB	5883	GGCCAGTGCATATGCAAGCAGAGGTGATCATGGGGGTGACAGTGTGACCCGTGACACCT	5942
QY	1601	ProLeuGlyPheGlyGlyLySerSerCyAlaGlnGlyMetAlaAsnProGlnHISPheLeu	1620	QY	1961	PheAlaGlyValThrThrAsnGlyCySerGlnValAsnTyrAspSerCyProArgAlaIle	1980
DB	4863	CCCCCTGGCTTTGGGGGCAAGAGCTGTGCCCCAGAAATGGCCAAATCCACACACTTCTG	4922	DB	5943	TTTGTAGGTGACCAACAAATGGCTGTGAAGTAAATATGACAGCTGCCACAGGCATTT	6002
QY	1621	GlySerSerLeuValAlaTyrHISGlyLeuSerLeuProIleSerGlnProTyrTyrLeu	1640	QY	1981	GlyAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCyPro	2000
DB	4923	GGCAGACCTCGTGGCTGGCAATGGCTCTGCTGCCATCTCCCAACCTTGTACTTC	4982	DB	6003	GAGGCTGGATCTGTGTGGCCCCGTAACCCGCTTGGGCTGCTGTGCTCTCCCTGCC	6062
QY	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660	QY	2001	LySGlySerPheGlyThrAlaValArgHISCyAspGlnHISArgGlyTyrPleuProPro	2020
DB	4983	AGCTCATGTTCCGCAACGCGCAGCGCGAGCGGTGTGCTGACAGGCATCACAGGGGG	5042	DB	6063	AAAGGCTCTTTGGAGACTGCTGTGGCCACTGTGATGACACAGGGGTGGCTCCCCCA	6122
QY	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyHISValMetLeuSerValGlnGlyThr	1680	QY	2021	AsnLeuPheAsnCySerThrSerIleThrPheSerGlnLeuLySGlyPheAlaGlyArgLeu	2040
DB	5043	CGCAGCACCATCACCCTACAGCTACAGAGAGGCCACGTGATGTGAGGTGAGAGGCCA	5102	DB	6123	AACTCTTCACTGACAGTCACTCACTTCTCAGAACTGAAGGGCTTGCAGAGCGGCTA	6182
QY	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTyrHIS	1700	QY	2041	GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg	2060
DB	5103	GGGCTTCAAGGCTCTCTCTCTCGTGTGAGGCCAGGCCCGGCAATGACGGTGACTGGAC	5162	DB	6183	CAGCGAAATAGTCAAGGCTAGACTCAGAGGCGCTCCAGAGAGCTAGGCTCTCTGCGC	6242
QY	1701	HISAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHISAlaAlaIleLeuSerPheAsp	1720	QY	2061	AsnAlaThrGlnHISThrAlaGlyTyrPheGlySerAspValIlyValAlaTyrGlnLeu	2080
DB	5163	CATGCACAGCTGGCACTGGAGACGAGCGGGGGCTTGGCAATGCCATCTCTCTTCAT	5222	DB	6243	AAAGCCACGACGACACAGCTGGCTACTTCCGACAGCGACTCAAGGTGGGCTTACACTG	6302
QY	1721	TyrGlyGlnGlnAlaGlyAlaGlyValAsnLeuGlyProAlaGlyMetHISGlyLeuHISLeuSer	1740	QY	2081	AlaThrArgLeuLeuAlaHISGlySerThrGlnAlaArgGlyPheGlyLeuSerAlaThrGln	2100
DB	5223	TATGGGCGAGAGAGCGAGAGGCCAACCTGGGCCCCCGGCTGACATGTGCTCACCTGAC	5282	DB	6303	GCCAGCGGCTGTGGGCCCAAGAGACCCACAGGGGGCTTGGGCTGTGTCACACAG	6362
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAspGlyPheArgGly	1760	QY	2101	AspValHISpHeThrGlnLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
DB	5283	AACTAACAGTGGGGGGAATACCTTGGGCCAGCGCGGTGTGGCCCTTTCGGGGGC	5342	DB	6363	GAGCTGACCTTCACTAGAGATCTGCTCGGGTGGGACGCCCTCTCTGACACAGCCAC	6422
QY	1761	CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer	1780	QY	2121	LySerGlySTrPglLeuLeuIleGlnGlnThrGlnGlyGlyThrAlaTyrPleuLeuGlnHIS	2140
DB	5343	TGTTTGCAGGGGTGTGCGGGTGAAGCATCGCCAGAGGGGGTTAACAGCTGATCCGACG	5402	DB	6423	AAAGCGCACTGGAGGCTGATCCAGACAGAGAGGTGGACCCGCTGGCTTCACAGAC	6482
QY	1781	HISGlyGlySerIleAsnValGlnGlnGlyCySerSerLeuProAspProCyAspSerAsn	1800	QY	2141	TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHISpHeThrTyrLeuSerProPhe	2160
DB	5403	CATGGGAGAGCATCAAGGTGAGACAGAGCTGTAGCTGTGACCTTGTGACCTCAAC	5462	DB	6483	TATGAGGCTTACGCAAGTGCCTGGCCCAAGACATGGGGCACTACCTTAAGGCCCTTC	6542
QY	1801	ProCySPProAlaAsnSerTyrCySerAsnAspTyrAspSerTyrSerCySerCyAsp	1820	QY	2161	ThrIleValThrProAsnIleValIleSerValValArgLeuAspLySGlyAspPheAla	2180
DB	5463	CGGTGTCTGTACAGCTATTGACAGAACAGACTGGGACACTATCTCTGACGCTGTGAT	5522	DB	6543	ACCATGTGACGGCCAAATGTGATCTCGTATGTGGCTTGGACAAAGGAATCTTGCT	6602
				QY	2181	GlyAlaLyLeuProArgTyrGlyAlaLeuArgGlyGlnGlnProProAspLeuGlnThr	2200



Db 6603 GGGGCGCAAGCTGGCCGCTACAGAGCCCTGGCGGAGACAGCCCGGAGCTTGAGACA 6662  
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
Db 6663 ACGATATCTGCTGAGTCTGCTTCACAGAGACCCCGCTGGTCAGGCCCGGAGGCG 6722  
Qy 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeu 2240  
Db 6723 CCGGAGAGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6782  
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db 6783 AGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6842  
Qy 2261 HisAsnTyrAspProAlaArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
Db 6843 CATACCTAGACCTTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6902  
Qy 2281 ProValAlaSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLys 2300  
Db 6903 CCGGTGTAGAGCATCAGGCTCCATGATGATGAGAGCTTCTGCGCGGCGCTGAGCAAA 6962  
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGlnGluArgThrLysProIleCysVal 2320  
Db 6963 CCGGTACGAGTGTCCGCTGCTGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAG 7022  
Qy 2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrSerAlaArgGlyCysVal 2340  
Db 7023 TTCTGGAACCATCAATCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7082  
Qy 2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
Db 7083 GTGCTCTTCGCGAATGAGAGGCAAGTCAAGTGCAGAGGCAACCAATGAGAGCTTGCT 7142  
Qy 2361 ValLeuMetAspValSerArgGluAsnGlyGluIleLeuProLeuLysThrLeuThr 2380  
Db 7143 GTGCTCATGAGCATTTCTCGGCGGAGAGATGAGGAGATCTGTCACCTGCAAGACATGACA 7202  
Qy 2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeu 2400  
Db 7203 TAGGTGCTTACAGTGTCACTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7262  
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
Db 7263 TTGCGATCTCTGCGCTCAACCAACAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 7322  
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
Db 7323 CTGGCTCAGCTGAGTCTTCTCTGAGGAGATCAACAGGCTGACCTCCCTTTGCTGACA 7382  
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGly 2460  
Db 7383 GTCATTCGCCATCTGCTGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 7442  
Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480  
Db 7443 GCGTTGACCTGTACCGGAGCATCACTGAGGTGCGAGTCAACACCGGCGCCCATAGCGG 7502  
Qy 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db 7503 TTCTACTACATGCTGGGCTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7562  
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuLeu 2520  
Db 7563 GACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7622  
Qy 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
Db 7623 TGGAGTTTCTGCTGCGCGGCTTGTGCGTCTGATGAGTGTCTTCCGTGATCACTCG 7682  
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyLysLysGlyProValSer 2560

Db 7683 GCGGCGCGGCGCTCTGCTGCTGCCAGCGGAGGAGGCTTTGAGAGAAAGTCTCTGCTCG 7742  
Qy 2561 GilyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrThrLeuLeuAlaLeu 2580  
Db 7743 GCGCTGCAAGCTCTCTTGGCGCTCTCTGCTGCTGAGCGGCACTGCTGCTGCTGCTG 7802  
Qy 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
Db 7803 CTTCTGTCAACAGAGACACCTCTCTTCCACTACCTCTTTGCTACCTGGAATTGCAATC 7862  
Qy 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlnValArgLysAlaLeu 2620  
Db 7863 CAGGCGCCCTTCATCTTCTCTCTATGTGTGCTTACAGAGAGTCCGGAAGACACTTC 7922  
Qy 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
Db 7923 AAGCTTGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7982  
Qy 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db 7983 ACTGTCTTCAACTGCGCCAGCGCTTACGAGATGGCGGCTGTACAGCCCTTACGGA 8042  
Qy 2661 AspSerAlaGlySerLeuHisSerThrArgSerGlyLysSerGlnProSerTyrIle 2680  
Db 8043 GACTGGCGGCTCTCTGCAACAGCACAGTGTGCTGCGGCAAGATCAAGCCAGCTTACATC 8102  
Qy 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db 8103 CCTTCTTGCTGAGGAGAGAGTCCGACTGAACTTGCCAGAGGAGGAGGAGGAGGAGGAG 8162  
Qy 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8163 GATCCAGGAGGAGCTGTCTTCTGAGAGTCAAGACAGAGCATGATCTGACAGGAGCTCC 8222  
Qy 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
Db 8223 GACAGTGAACCTGTCTTGAAGACCAAGAGTGTCTCTTATGCTTACCACTCATCA 8282  
Qy 2741 AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
Db 8283 GACAGTGAAG 8342  
Qy 2761 TrpAspSerLeuLeuGlyProGlyValAlaGluArgLeuProLeuHisSerThrProLysAsp 2780  
Db 8343 TGGATAGGCTGCTGCGGCTGAGACAGAGAGCTGCCCTGCAAGTACTCCCAAGAT 8402  
Qy 2781 GlyGlyProGlyProGlyLysAlaProThrProGlyAspPheGlyThrThrAlaLysGlu 2800  
Db 8403 GGGGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8462  
Qy 2801 SerSerGlyAsnGlyValaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
Db 8463 AGTATGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8522  
Qy 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyIleLeuLys 2840  
Db 8523 GAGGGGTCTCTAGGCGCCCTTCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 8582  
Qy 2841 LysLysCysLeuProThrIleSerGlyLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
Db 8583 AAGAGGTGTCTGCCCACTACAGCAAGAGAGACACTTCTGCGCTCCCTGAGAGCAA 8642  
Qy 2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
Db 8643 TGCAAGGGTCTTCCCGGGGCTCTCCGCTATGAGGGGAGCGGAGGAGGAGGAGGAGGAG 8702  
Qy 2881 ArgProProProArgGlnSerLeuGlnGlnLeuAsnGlyValMetProIleAlaMet 2900  
Db 8703 CCGCCACGCGCCCGGAGAGGCTTCCAGAGCAGCTGAACAGGGGTATGCTTCCGCAAG 8762  
Qy 2901 SerIleLysAlaGlyThrValAspGluAspSerSerGlySerGlyPheLeuPhePheAsn 2920  
Db 8763 AGCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8822

OY 2921 Pheleuhis 2923  
 DB 8823 TTCTGAT 8831

## RESULT 3

ABX34546  
 ID ABX34546 standard; cDNA, 11762 BP.

XX ABX34546;

XX 13-FEB-2003 (first entry)

XX Human mdct cDNA SEQ ID 107.

XX MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianemic; antiproliferative; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;  
 KW psoriasis; hepatitis; gene; ss.

XX Homo sapiens.

XX WO200279449-A2.

XX 10-OCT-2002.

XX 27-MAR-2002; 2002WO-US09944.

XX 28-MAR-2001; 2001US-279619P.

XX 29-MAR-2001; 2001US-280067P.

XX 29-MAR-2001; 2001US-280068P.

XX 16-MAY-2001; 2001US-291280P.

XX 17-MAY-2001; 2001US-291829P.

XX 17-MAY-2001; 2001US-291849P.

XX 19-JUN-2001; 2001US-299428P.

XX 20-JUN-2001; 2001US-299776P.

XX 20-JUN-2001; 2001US-300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Dafio A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PB, Amshy SR;

XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;

XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-058431/05.

XX P-PSDB; ABU11556.

XX New purified disease detection and treatment molecule proteins and

XX polynucleotides, useful for diagnosing, treating or preventing cancers

XX (e.g. leukemia or sarcoma), anemia, Crohn's disease, osteoporosis

XX or hepatitis

Claim 1; SEQ ID NO 107; 339pp + Sequence Listing; English.

This invention describes a novel disease detection and treatment molecule polypeptide (MDDT) which has anti-inflammatory, immunosuppressive, osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic, antianemic, antiproliferative and hepatotropic activity. The polynucleotides and the polypeptides of the invention can be used for gene therapy, protein replacement therapy and are useful for treating a variety of diseases or conditions. These polypeptides or polynucleotides are particularly useful for diagnosing, treating or preventing cell proliferative disorders (e.g. cancers including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's syndrome, inflammation, osteoporosis, thrombocytopenia, psoriasis or hepatitis. ABX34440-ABX34835 encode the MDDT polypeptides, represented in

CC ABU11450-ABU11845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 11762 BP; 2286 A; 3738 C; 3283 G; 2455 T; 0 other;

## Alignment Scores:

Pred. NO.:	0	Length:	11762
Score:	2894.00	Matches:	2894
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.01%	Indels:	0
DB:	25	Gaps:	0

US-09-916-849a-3 (1-2923) x ABX34546 (1-11762)

OY	30	LeuGIYAspGlnValGlyProCyAsrSerLeuGIYSerArgIYValGlySerSerGIY	49
DB	2	TTGGGAGACCAAGTGGGGCTGTCTTGGGGTCCAGGGGACGAGGCTTTCGGGG	61
OY	50	AlaCyAspAlaProMetGIYTPleuCyProSerSerAlaSerMetLeuTPleuTYThr	69
DB	62	GCTGGCGCCCAATGGGCTGTCTTCTTCATCTTCAGCGTCAACCTGTGCTTACACC	121
OY	70	SerArgCyAsrAspAlaGIYThrGluLeuThrGIYHisLeuValProHisAspGIY	89
DB	122	AGCGCTGACGAGGATGGGGCACTGAGCTGACCTGGCACTTCCACGACGATGGC	181
OY	90	LeuArgValITProCyProGluSerGIYValHisITLeProLeuProAlaProGluGIY	109
DB	182	CTGAGGGTGTGTGTCCAGAAATCCAGGCCCATATTCCTCCACACGACTCTGAAAGGC	241
OY	110	CysProITProSerCyAsrArgLeuGIYITLeuGIYHisITLeuSerProGluGIYLeu	129
DB	242	TGCCCTGGAGCTGTGCTCTCTGGGATGAGGCCACTTCTCCACAGGGCAAGCTC	301
OY	130	ThrLeuProGluGIYHisITProCyAsrLeuValAlaProArgLeuArgCyGlnSerCyAs	149
DB	302	ACACTGCCGAGGAGGACCCGTGCTTAAGGCTCCAGGCTCAGATCCAGTCTGCAAG	361
OY	150	LeuValAGlnAlaProGluLeuArgAlaGIYValArgSerProGluLeuGIYVal	169
DB	362	CTGGCACAGGCCCCCGGGCTCCAGGGCAAGGATCCACAGAGGCTCTGGGGGGG	421
OY	170	ArgArgLeuArgAsnValAsnThrAlaProGlnPhleGluProProSerTYrGlnAlaThr	189
DB	422	CGTGGAAAGAAATGTAATTAACAGCCCCAGTTCAGGCCCCCGAGCTACAGGCCACA	481
OY	190	ValProGluAsnGlnProAlaGIYThrProValAlaSerLeuArgAlaITLeuProAsp	209
DB	482	GTGCCGAGAACCAAGCAGAGGACACCCGTGTGATCCCTGAGGGGCATCGACCGGAC	541
OY	210	GluGluGluValAGlValArgLeuGIYTYrMetAspAlaLeuPhleAsrArgSerAsn	229
DB	542	GAGGGTGAAGCAGTGTGACAGTACACCATGATGAGGCCCTTGTGATACCCCTCCAC	601
OY	230	GlnPhleSerLeuAspProValIThrGluValAlaThrAlaGluGluLeuAspArg	249
DB	602	CAAGTCTTCTCTCCAGAACCAAGTCACTGAGTCAAGTCAACACAGCCAGAGCTGAGTGT	661
OY	250	GluIThrLeuSerThrHisValPheArgValIThrAlaGlnAspHisITLeuMetProArg	269
DB	662	GAGACCAAGGACCAACGCTTCAAGGTCAAGGCCACAGGACCAAGGCCAGCCGACGA	721
OY	270	SerAlaLeuAlaThrLeuThrITLeuValIThrAspThrAsnAspHisAspProValPhe	289
DB	722	AGTGCCTGTGCTACACTCCATCTGTGTACGACCAACATACATGACCTGTGTTTC	781
OY	290	GluGluGluGluITyrITyrGluSerLeuArgValAsnLeuGluValGlyTYrGluValLeu	309
DB	782	GAGCAGCAGAGTCAAGAGAGAGCTCAGGAGAACTTGAGGTGGCTATGAGGTGCTC	841

QY 310 ThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeu 329  
DB 842 ACTGTGAGGGCCACGGATGGTGAATGCCCTCCCAATGCCAAATTCCTGATACGGCTGCTG 901  
QY 330 GluGlySerGlyGlySerProSerGlyValPheGluIleAspProArgSerGlyValIle 349  
DB 902 GAGGGGTCTGGGGGAGCCCTCTGAAAGCTTTGAGATGACCCCTGCTCTGGGGGAATC 961  
QY 350 ArgThrArgGlyProValAspArgGluGluValGlySerTyrGlnLeuThrValGluAla 369  
DB 962 CGAACCCGTGGCCCTGTGATCGGGAAGGTGGAAATCTTACACGCTGACGGTAGAGGA 1021  
QY 370 SerAspGlnGlyArgAspProGlyProArgSerThrThrAlaAlaPheLeuSerVal 389  
DB 1022 AGTGACACAGGGTCTGGGACCCGGGTCTCGGAGTACACAGCCGCTGTTTCTCTTCTG 1081  
QY 390 GluAspAspAsnAspAsnAlaProGlnPheSerGlyLeuArgTyrValValGlnValArg 409  
DB 1082 GAGGATGACAAATGATATGCCCCCAGTTTATGAGAGACGGCTATGTGTCCAGGTGAG 1141  
QY 410 GluAspValThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspGly 429  
DB 1142 GAGGATGATCTCGAGGGGCCCAAGTACTCCAGTCAAGCTCCGATCGAGACAAAGGG 1201  
QY 430 SerAsnAlaValAlaHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeu 449  
DB 1202 AGCAATGCCGTGGTGCATATAGCATCATGAGTGCATGTCTGGGGCAGATTATCTG 1261  
QY 450 AspAlaGlnThrGlyValLeuAspValValSerProLeuAspTyrGlnThrTyrLeu 469  
DB 1262 GATGCCCAATCGAGACTGTGATGTGTGAGCCCTTGACTATGAGCGACAAAGAG 1321  
QY 470 TyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGly 489  
DB 1322 TACACCTTACGGGTGGAGACACAGATGGTGGCCGCCCTCCACTCTTATGTCTTGGC 1381  
QY 490 LeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrPro 509  
DB 1382 TTGGTGAACAGTACAGGTCTCGATATCAACAGATGCCCATCTTCTCTGACACCCCT 1441  
QY 510 PheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAla 529  
DB 1442 TTCAGGCTACTGTCTCTGAGAGCGTCCCTTATGGCTGACTGTCTCTCATGTCCAGGCT 1501  
QY 530 IleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTyrArgLeuAlaGlyValGlyHis 549  
DB 1502 ATCGACGCTGATGCTGTGACAAATGCCGCTGGAAATCCGCTTCTGTGGGTGGAGCAT 1561  
QY 550 AspPheProPheThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAsp 569  
DB 1562 GACTTCCCTTCAACATCAACATGACAGGCTGATCTGTGTGCTGTGAACCTGAGC 1621  
QY 570 ArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAla 589  
DB 1622 CGGAGAGAGTGAATTTCTACAGCTTTGGGGTAGAAGCTCGAGACCAATGCACTCAGGA 1681  
QY 590 LeuThrAlaSerAlaSerValSerValThrValLeuAspValAsnAspAsnProThr 609  
DB 1682 CTCACGTGCTCGGCGACGTGACGCGTACTGTCTGGATGTCACAGCAACAAATCAACC 1741  
QY 610 PheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerVal 629  
DB 1742 TTTTACCACCAACAGATGACAGTGGCGCTCAATGAGATGACGTGTGGCCACGCGTG 1801  
QY 630 ValThrValSerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSer 649  
DB 1802 GTGACCGGTCTCAGCTGTGACCGCTGATGCTCATAGTGTATCACTTACCAAGTACCAAT 1861  
QY 650 GlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeu 669  
DB 1862 GGCAATATCTCGAAACGGCTTCTCATCACACCAAGGTGTGGGTGTATCCCTT 1921  
QY 670 AlaLeuProLeuAspTyrIleLeuGluArgGlnTyrValIleValAlaThrAlaSerAsp 689

DB 1922 GCCCTGCCACTGACCTACAACTTGAGCGGACGATGTGTGGCTGTATACGGCTCCGAT 1981  
QY 690 GlyThrArgGluAspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArg 709  
DB 1982 GGCATCTGGAGGACACGGGACAGATGTGTGATGATGACCGAGCAACCACTCATGT 2041  
QY 710 ProValPheGlnSerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThr 729  
DB 2042 CCTGTCTTCAAGAGTCTCCCATATACGTGATGTATATGAGACCGGCCGAGGACAC 2101  
QY 730 ThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyr 749  
DB 2102 ACGGTGTGTGATACAGCGCCACGATGAGACACAGGTGAGAAATCCCGCATCACTAC 2161  
QY 750 PheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyValValThrThr 769  
DB 2162 TTCATGAGAGACAGATCCCGACGTTCCGATCCATGACATGACACAGGGGGCTGTACACCC 2221  
QY 770 GlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAsp 789  
DB 2222 CAGGTGAGCTGAGCTACGAGACCAAGTCTTACACCTTGCCATTACTGCTCGGAC 2281  
QY 790 AsnGlyIleProGlnLeuSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsn 809  
DB 2282 AATGCGATTCGCCAGAGTCCGACACCACTTACCTGAGATCTGTGTGAAACGAGTGAAT 2341  
QY 810 AspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValPro 829  
DB 2342 GACATATGCCCTCAGTTCTCTCGAGACTCTTACAGGGCAGATGTCTATGAGAGATGTCCA 2401  
QY 830 ProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArg 849  
DB 2402 CCTTCACTACGCTGCTGACATCTCAGCCACTGATGTGATTCCTGACCTTAATGGCAGG 2461  
QY 850 ValPheTyrThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThr 869  
DB 2462 GCTTCTTACACCTTCCAGAGGGGCGACATGACACGCTGATTTATTTGTGTGATCTCAGC 2521  
QY 870 SerGlyIleValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeu 889  
DB 2522 TCAGGATCTGTGCAAGCGCTACGAGGCTGATGAGAGAACGTGGCCAGTATGTCTTG 2581  
QY 890 ArgAlaTyrAlaValAspLeuGlyMetProProAlaArgThrPrometGluValThrVal 909  
DB 2582 CGGCAATATGACAGTGAACAGGGGATGCCCAACCCGCACTATGAGAGTGAACGCTC 2641  
QY 910 ThrValLeuAspValAsnAspAsnProProValPheGluGlnAspGluPheAspValPhe 929  
DB 2642 ACTGTGTGATGTGAAATGACAAATCCCTGTCTTTGAGCGAGATGAGTTGATGTGTTT 2701  
QY 930 ValGluGluAsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAsp 949  
DB 2702 GTGAGAGAGACACCCCATTTGGGCTAGCCGTGGCCGGGTCAAGCCATCAAGCCCGAT 2761  
QY 950 GluGlyThrAsnAlaGlnIleMetTyrGlnIleValGluGlyAsnIleProGluValPhe 969  
DB 2762 GAAAGCACCAATGCCAGATTAATGATACAGATGTGAGGGCAACATCCCTGAGGTCTTC 2821  
QY 970 GlnLeuAspIlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArg 989  
DB 2822 CAGCTGACATCTTCTCCGGGAGACTGACAGCCCTGTGTAGCTTATGACTACAGAGACCGG 2881  
QY 990 ProGluTyrValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrVal 1009  
DB 2882 CCTGAGTACGTCCTGCTGATCTACAGGCCACGTCACCTCTCTGTGTGAGCCGGGCTCAAGTCTC 2941  
QY 1010 HisValArgLeuLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeu 1029  
DB 2942 CACGTCCGCTCTCTGACCGGAAATGACAACTCCATGTGTGGGCACTTTGAGATCTCT 3001  
QY 1030 PheAsnMetTyrValThrAsnArgSerSerPheProGlyGlyAlaIleGlyArgVal 1049

Dp	3002	TTCAACAACATATGATCAACCAATGCTCAAGACAGCTTCCTGGGGGGTGCAATTGGCCGAATA	3061
Qy	1050	ProLah1AspProAsp11eserAspserLeuThr1YrSerPheGluAraG1YAsnGlu	1065
Dp	3062	CCTGCCATGACCTCGATATCTCAATATGTCTGACTTACGCTTTGAGCGGGGAATGAA	3121
Qy	1070	LeuSerLeuValLeuLeuAsnAlaSerThrg1Yg1LeuLysLeuSerArgAlaLeuAsp	1089
Dp	3122	CTGACCCCTGCTCTCTCAATCTCCATCCCTCCACGGGTGTGCTGAAGCTTAACCGCGCACTGGAC	3181
Qy	1090	AsnAsnArgProLeuGluAla11eMetSerValLeuValSerAspG1YAlH1SerSerVal	1109
Dp	3182	AACAAACCGGCTCTTGAGGCAATCAATGACGGCTGTGTCAAGCGGGTCAACAGCGCTG	3241
Qy	1110	ThralaG1NcySa1aLeuAraGValThr11e1ThrAspGluMetLeuThH1Ser11e	1129
Dp	3242	ACCGCCAGTGGGGCTGGTGTTGACATACATACCAAGTAGATCTCAACCAACGATC	3301
Qy	1130	ThrLeuAraGLeuGluAspMetSerProGluAraPheLeuSerProLeuLeuG1YLeuPhe	1149
Dp	3302	ACGGTCCGCTCGAGAGACATGTCAACCGAGCGCTTCTGTCAACCACTCTAGCGCTTTC	3361
Qy	1150	11eGln1aValAla1a1a1aThrLeuAlaThrProProAspH1SValValValPheAsnVal	1165
Dp	3362	ATCCAGCGGGGTGGCCGCAACGTGTGCACGCGCACCGAGCAACAGTGTGTCTTCAACCTA	3421
Qy	1170	GlnAraGAPThrAspAlaProG1Yg1YH1S11eLeuAsnValSerLeuSerValG1Yg1Gln	1189
Dp	3422	CAGCGGACACCGAGCGCCCGCGGGGGCCAACTCTCAACGTAGACTTCTGTGGTGGCCAG	3481
Qy	1190	ProProG1YProG1Yg1Yg1YProProPheLeuProSerGluAspLeuGlnG1YAraGLeu	1209
Dp	3482	CCGCGAGGGGCGGGGGGGGGGGCGGCGCTTCCGTCCGCTCTGAGAACCTCAAGAGGCGCTA	3541
Qy	1210	Tyr1LeuAsnAraGserLeuLeuThrAla11eSerAlaGlnAraGValLeuProPheAspAsp	1229
Dp	3542	TACCTCAACCGAGCTGTGTGACGGCCAACTCTGGACAGCGGGTGTCCCTTGCACACAC	3601
Qy	1230	Asn11eCysLeuAraG1YProCysG1YAsnAraYrme1aGcysValSerValLeuAraPhe	1249
Dp	3602	AACACTCTGCTGGCGGAGACCTCGCGAAGAACTACAGCCCTGTGTGTGTGTGTGTGTTC	3661
Qy	1250	AspSerSerAlaProPhe11eAlaSerSerSerValLeuPheAraGPro11eH1AsProVal	1265
Dp	3662	GACTCTCTCGCGGCTTCTCATGCGCTCTCTCCGCTCCGCTTCCGCGCCATCAACCCGCTC	3721
Qy	1270	G1Yg1YLeuAraGcysAraGcysProProG1YPheThrg1YAspYrYcysGlnThrg1Val	1289
Dp	3722	GAAGGGCTGTGGTGTGGCTGTGGCTCCCGCGGCTTCAACGGGTGTACTCTGTGAACCGAGGTG	3781
Qy	1290	AspLeuCyS1YrYSerArgProCysG1YProH1Sg1YArGcysValSerSerArgG1Yg1Yg1Y	1309
Dp	3782	GACTCTCTCTACTCCCGGCGCTGTGTGGCCCCCAAGGCGCGCTGTCCGACGCGGAGGGCGGG	3841
Qy	1310	TyrThrCysLeuCySArgAspG1YThrThrg1Yg1YH1S1Yg1YValSerAlaArgSer	1329
Dp	3842	TACACTGCTCTGTGTGTGATGGCTACACGGGTAGACCTGTGAGGTAGTGTGTGTGTCTCA	3901
Qy	1330	G1YArGcysThrProG1YValCysYbAraGlnG1YThrCysValAsnLeuLeuValG1Y	1349
Dp	3902	GGCCGCTTCAACCCCGGGGTGTGTGTGCAAGATGGGGCACTGTGTCAACTCTGTGTGGG	3961
Qy	1350	G1YPheLysCysAspCysProSerG1YAspPheG1YAspProYrYcysGlnVal1ThrThr	1369
Dp	3962	GGTTTCAAGTGTGATGGCCAACTGTGAAGCTTCAAGAAAGCCCTTATGTCAAGGTATCAACG	4021
Qy	1370	ArgSerPheProAlaH1SerPhe11eThrPheAraG1YLeuAraGlnAraPheH1SPh	1389
Dp	4022	CGAGCTTCCCCGAGCACTCTTATCACTTTGTGGCGGCTGTGGCAACGTTTCAACTTC	4081
Qy	1390	ThrLeuAlaLeuSerPheAlaThr1Yg1YArGAspG1YLeuLeuLeuYrYAsnG1YArG	1409
Dp	4082	AACCTGGGCTCTGTGTGTGTGCAAAAGAGCGCAAGGGTGTGTGTGTGTGTGTGTGTGTGTGT	4141

[illegible]

QY 1770 ThrProGluGlyValLeuSerLeuAspProSerHisGlyLeuSerLeuAsnValGluGln 1789  
DB 5222 AGCCCAAGAGGGGGTTAAAGCCCTGGATCCAGCCAGGGGAGGCAATCAACGGAGCAAA 5281  
QY 1790 GlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSer 1809  
DB 5282 GGGCTGAGCTGGCTTACCTCTGAGCTCAAAACCCGTCGCTGCTAAACAGCTATTTGACG 5341  
QY 1810 AsnAspTrpAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyValAsnAsnCysThr 1829  
DB 5342 AACGACTGGGAGAGCTATTCCTGCACTGATGATCCAGGTTACTATGCTGCAACTGTACT 5401  
QY 1830 AsnValCysAspLeuAsnProCysGluHisGlnSerValCysThrArgLysProSerAla 1849  
DB 5402 AATGTGTGTGACCTGTAACCCGTGTGTGACCACTGTGTGTATCCCGCAAGCCCACTGACC 5461  
QY 1850 ProHisGlyTyrThrCysArgLysCysProProAsnTyrLeuGlyProTyrCysGlyIleThrArg 1869  
DB 5462 CCCCATGGCTATTCCTGCAAGTGTCCCCCAATTAACCTTGGGCACTATCTGTAGACCAAG 5521  
QY 1870 IleAspGlnProCysArgProArgGlyTyrTrpGlyHisProThrCysGlyProCysAsnCys 1889  
DB 5522 ATTGAACAAGCTTTGTCCCGGTGGCTGTGGGAGCATCCCAATGTGGCCCATGCACTGT 5581  
QY 1890 AspValSerLysArgLysPheAspProAspCysValAsnLysThrSerGlyLysCysHisCysLys 1909  
DB 5582 GATGTGAGCAAAAGCTTTGACCACTGCACTGCAACAAAGCCGCGCAAGTGTCCACTGCAAG 5641  
QY 1910 GluAsnHisGlyArgProProGlySerProThrCysLeuLeuCysAspCysTyrTrpProThr 1929  
DB 5642 GAGAACCACTACCGGCCCCCAAGCAGCCCACTGCTCTTGTGTATCTGTACCCCAAA 5701  
QY 1930 GlySerLeuSerArgValCysAspProGlyAspGlyGlnCysArgProCysLysProGlyVal 1949  
DB 5702 GGCCTCTTGTCCAGAGCTGTGTGACCTGAGGAGTGCAGGTGTCAATGCAAGCAGGTGTTC 5761  
QY 1950 IleGlyArgGlnCysAspArgCysAspAsnProPheAlaGlyValIleThrThrAsnGlyCys 1969  
DB 5762 ATCGGGCGTCAAGTGTACCGGTGTGACAAACCTTTGTGTGAGGTGCAACCAATGGCTGT 5821  
QY 1970 GluValAsnTyrAspSerCysProArgAlaIleGlyAlaGlyIleTrpTrpProArgThr 1989  
DB 5822 GAAAGTAATTATTCAGCTGTGCCAAGAGCATTTGAGGCTGGATCTGTGTGCCCCCTAAC 5881  
QY 1990 ArgPheGlyLeuProAlaAlaAlaProCysProLysGlySerPheGlyThrAlaValArg 2009  
DB 5882 CCTTCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5941  
QY 2010 HisCysAspGlnHisArgGlyTyrLeuProProAsnLeuPheAsnCysThrSerIleThr 2029  
DB 5942 CACTGTGATGAGCAAGGGGGTGGCTCCCCCAAACTTTCACATGCAAGTCCATCACCC 6001  
QY 2030 PheSerGlyLeuLysGlyPheAlaGlyLysGlyLeuGlnArgAsnGlySerGlyLeuAspSer 2049  
DB 6002 TTCTCAGAACTGAAGGGCTTCCGTGAGCGGCTTACACGGAACTAGAGTCAAGGCCCTAGCTCA 6061  
QY 2050 GlyArgSerGlnGlnLeuAlaLeuLeuLeuAsnAlaThrGlnHisIleThrAlaGlyTyr 2069  
DB 6062 GGGCGCTCCCAAGCACTAGCCCTGCTCTGCGCAAGCCCAAGCAGCAAGCACTGGCTAGC 6121  
QY 2070 PheGlySerAspValLysValAlaTyrGlnIleValIleThrArgLeuLeuAlaHisGlySer 2089  
DB 6122 TTGGGAGGAGCACTGAAGGTGGCTTACCAAGTGGCCCAAGGGGCTGGCCCAAGAGG 6181  
QY 2090 ThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisIlePheThrGlyIleAsnLeuLeu 2109  
DB 6182 ACCCAAGGGGGCTTTGGGCTGTGTGCAACAAGACAGTGTGCACTTCACTAGAACTGCTGG 6241  
QY 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisIleTrpGlyLeuIleGlnGln 2129  
DB 6242 CGGGTGGGAGCGCCCTCTGCAACAGCCCAAGCGGCACTGGAGCTGAACTCCAGCAG 6301  
QY 2130 ThrGlnGlyLysIleThrAlaTrpLeuLeuGlnHisGlyGlyIleValIleThrAlaSerAlaLeuAla 2149

DB 6302 ACAGAGGGTGGCAACCGCTGTGCTTCCAGCACTATGAGGCTTACGCAAGTCCCTGCGCC 6361  
QY 2150 GluAsnMetCysArgHisIleThrTyrLeuSerProPheThrIleValIleThrProAsnIleValIle 2169  
DB 6362 CAGAACATTCGGGCAACCTTACCTTAAGCCCTTCAACATCTGTACAGCCCAACTTGTCTATC 6421  
QY 2170 SerValValArgLeuAspLysGlyAsnPheAlaGlyAlaLysLeuProArgTrpGlyVal 2189  
DB 6422 TCCGTAGTGGCTTGTGCAAAAGGAACTTTGTGTGGGCAAGAGTCCCGCTGACAGAGGCC 6481  
QY 2190 LeuArgGlyGlyGlnProProAspLeuGlyIleThrThrValIleLeuProGluSerValPhe 2209  
DB 6482 CTGCGTGGGAGAGCAACCCCGGACCTTGTAGCAACAGTCATTCTGCTGAGCTGTGTCTTC 6541  
QY 2210 ArgGlyIleThrProProValValArgProAlaGlyProGlyGlyAlaGlnIleProGlyGly 2229  
DB 6542 AGAGAGAGCCCCCGCTGTGTACAGCCCGGAGCTGAGCCAGGAGAGGCTGTGGCCAGCGTC 6601  
QY 2230 LeuAlaArgArgGlnArgArgHisProGlyLeuSerGlnGlyAlaValAlaLeuSerVal 2249  
DB 6602 CTGGCAACGGCAACAGCAAGGCAACCGGAGCTGAGCCAGGAGTGTGGCCAGCGTC 6661  
QY 2250 IleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProAspLysArgSer 2269  
DB 6662 ATCATCTACCGCAACCTGCGCGGCTACTGCTCTCAATTAACCTGACCAAGCGCAGC 6721  
QY 2270 LeuArgValProLysArgProIleIleAsnThrProValIleSerValHisAsp 2289  
DB 6722 TTGAGAGTCCCAAAACGCCCATCATCAACACCCGTGTGAGATCAGCTCCATGAT 6781  
QY 2290 AspGlyIleLeuLeuProArgAlaLeuAspLysProValIleThrValGlnPheArgLeuLeu 2309  
DB 6782 GATGAGAGAGCTTTCGCCCGGCGCTGGAACAACCCGTGCAAGTGTCACTCCGCTGCTG 6841  
QY 2310 GlyThrGlyGlyLysArgThrLysProIleCysValPheTrpAsnHisSerIleLeuValSer 2329  
DB 6842 GAGACAGAGAGCGGACCAAGGCCATCTGTGTCTTGTGACCAATTCATCTGTGATCAGT 6901  
QY 2330 GlyThrGlyGlyTyrPheSerAlaArgGlyCysGlyValValPheArgAsnGlySerHisVal 2349  
DB 6902 GGCACAGGTGGCTGTGCGGCAAGAGCTGTGAAGTGTCTTCCGCAATGAGAGCCACGTTC 6961  
QY 2350 SerCysGlnCysValAsnHisMetThrSerPheAlaValLeuMetAspValSerArgArgGly 2369  
DB 6962 AGCTCCAGTGTGAACACATGACAGAGCTTCCGTGTGTGATGAGAGTTCCTCGGGGAG 7021  
QY 2370 AsnGlyGlyIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValIleLeuAla 2389  
DB 7022 AATGGGAGATCTCTGCACTGAAAGCACTGAAATAGTGGCTTATGATGTCACTTGGCT 7081  
QY 2390 AlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHis 2409  
DB 7082 GCCCTTGTGCTCACTTCTTCTTCTCACTGCTTGGATCTGCGCTCCCAACCAAC 7141  
QY 2410 GlyIleLeuArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGly 2429  
DB 7142 GGCATCCCAAGTAACTAAGCTGCGCTGGGCTGCTGAGCTGTCTCTCTGAG 7201  
QY 2430 IleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHisIlePheLeu 2449  
DB 7202 ATCAACAGGCTGACCTTCTTGTGCTGCAAGTATGCAATTCCTGCTGCACTTCTG 7261  
QY 2450 TyrLeuCysThrPheSerTrpAlaLeuLeuGlyAlaLeuHisIleLeuTyrArgAlaLeuThr 2469  
DB 7262 TACCTGTGACCTTTTCCGGGCTGTGAGGAGCTTGCACCTGTACCGGGCACTCACT 7321  
QY 2470 GluValArgAspValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrTrpVal 2489  
DB 7322 GAGGTGGGATGTAAACCGGCCCCAAGGCTTCTACTATCTGAGCTGGGAGCGTG 7381  
QY 2490 ProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGlyGlyTyrGlyAsnProAsp 2509

Db 7382 CCTGCTTCATCAGAGGCTGAGCCGTGGGCTGAGACCCCGAGGCTACAGGACCTTAC 7441  
 Qy PheCysTrpLeuSerLeuTyrAspThrLeuIleTpsSerPheAlaGlyProValAlaPhe 2529  
 Db 7442 TTCTGCTGCTCTCATCATATGACAGCTCATCTGAGATTTGGCTGGCCGGTGGCTTT 7501  
 Qy 2530 AlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGln 2549  
 Db 7502 GCCGCTCGATGAGTGTCTTCTCTGATCATCTCGGCGGCGCCGGCTCTCTGTGCTCCAG 7561  
 Qy 2550 ArgGlnGlyPheGlnLysGlyProValSerGlyLeuGlnProSerPheAlaValLeu 2569  
 Db 7562 CGGACAGGCTTGGAGAAAGATCTGTCTCGGGCTGAGCCCTCTCTCGCGCTCTC 7621  
 Qy 2570 LeuLeuSerAlaThrTrpLeuLeuAlaLeuSerValAspSerPheThrLeuLeu 2589  
 Db 7622 CTGCTGCTGAGGCGACCGTGGCTGGGACTGCTCTGTCTGTAACAGAGACCCCTCTC 7681  
 Qy 2590 PheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyr 2609  
 Db 7682 TTCCACTACCTCTTGTCTACCTGCAATTGCATCCAGGCGCCCTTCATCTCTCTCAT 7741  
 Qy 2610 ValValLeuSerLysGlnValArgValAlaLeuLysLeuAlaCysSerArgLysProSer 2629  
 Db 7742 GTGGTGTCTTACCAAGAGGTCGGAAGCACTCAAGCTTGGCTGACGCGCAAGCCAGC 7801  
 Qy 2630 ProbAbpProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsnCysProSerPro 2649  
 Db 7802 CCTACCTCTGCTGACCAACAGGTCACCCCTGACCTGTCTTCACTGCCCCAGCCCC 7861  
 Qy 2650 TyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThr 2669  
 Db 7862 TACCGAGATGGGGCGGCTTACCAAGCCCTACCGAGCTCGGCGGCTCTCTGACAGCAC 7921  
 Qy 2670 SerArgSerGlyLysSerGlnProSerTyrIleProPheLeuLeuArgGlnLysSerAla 2689  
 Db 7922 AGTCGCTCGGCAAGATCAAGCCAGCTACATCCCTCTTGTGAGGAGAGATCCGCA 7981  
 Qy 2690 LeuAbnProGlyGlnGlyProProGlyLysLeuGlyAspProGlySerLeuPheLeuGlnGly 2709  
 Db 7982 CTGAACCTTGGCCAGAGGCGCCCTGGGCTGGGGAGTCCAGGAGCGCTGTCTCTGAGAGGT 8041  
 Qy 2710 GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAsp 2729  
 Db 8042 CAAGCCAGCAGCATGATCTGACACGAGCTCCGACATGACTGTCTTGAAGAAGAC 8101  
 Qy 2730 GlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnGln 2749  
 Db 8102 CAGAGTGGCTCTTATGCTCTTACCCATCATCAGACAGTGAAGAGAGAGAGAGAG 8161  
 Qy 2750 GlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTrpAspSerLeuLeuGlyProGlyAla 2769  
 Db 8162 GAAAGAGAGGCGCTTCTCTGAGAGAGGAGGAGGAGTACCTGTGGGAGCTGAGAGCA 8221  
 Qy 2770 GlnArgLeuProLeuHisSerThrProLysAspGlyGlyProGlyLysAlaPro 2789  
 Db 8222 GAGAGACTGCGCCCTGACAGTACTCCCAAGATGGGGGCCCAAGGCGCTGCAAGGCCCC 8281  
 Qy 2790 TrpProGlyAspPheGlyThrThrAlaLysGlnLysSerGlyLysGlnAlaProGlnGln 2809  
 Db 8282 TGGCCAGAGAGACTTGGAGCAAGCAAAAGATGATGCAACGAGGCGCCCTGAGAGAG 8341  
 Qy 2810 ArgLeuArgGlnGlnGlnAlaLeuSerArgGlnGlnLysSerLeuGlnProLeuProGly 2829  
 Db 8342 CGGCTGCGGAGAAAGATGCTGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8401  
 Qy 2830 SerSerAlaGlnProHisLysGlyLysLeuLysLysCysLeuProThrIleSerGln 2849  
 Db 8402 TCTTGTGCGCAGGCTCAAAAGGATCTTAAAGAAAGATGTCTGCCACATCAGAGAG 8461  
 Qy 2850 LysSerSerLeuLeuArgLeuProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2869  
 Db 8462 AAGAGCAGCTCTGCGGCTCTCTGAGAGCAATGACAGGAGTCTTCCCGGAGCTCTCTC 8521

Qy 2870 AlaSerGlnGlySerArgGlyGlyProProProAbpProProProArgGlnSerLeuGln 2889  
 Db 8522 GCTAGTAGAGGAGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8581  
 Qy 2890 GlnGlnLeuAbnGlnValMetProIleAlaMetSerIleLysAlaGlyThrValAspGln 2909  
 Db 8582 GAGCAGCTGAACGGGCGATGCGCATGCGCATGAGATCAAGAGCAGGAGCGTGGATGAG 8641  
 Qy 2910 AspSerSerGlySerGlnPheLeuPhePheAsnPheLeuHis 2923  
 Db 8642 GACTCGTCAAGCTCCGAATTTCTTCTTAACCTTCTGAT 8683  
 RESULT 4  
 ID AS11677 standard; cDNA; 8871 BP.  
 AC AS11677;  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human Flamingo cDNA sequence.  
 XX  
 KW Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
 KW signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
 KW obesity; hypertension; hyperextension; urinary retention; angina pectoris;  
 KW myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
 KW benign prostatic hyperplasia; psychotic disorder; neurological disorder;  
 KW Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KW antifungal; antiviral; antiparasitic; anti-HIV; anorectic; antianorectic;  
 KW antiparkinsonian; cardiac; cerebroprotective; neuroprotective;  
 KW antidepressant; anticonvulsant; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1..8771  
 FT /tag= a  
 FT /product= "Human Flamingo protein #1"  
 XX  
 XX W0200161003-A1.  
 XX PD 23-AUG-2001.  
 XX PP 19-FEB-2001; 2001M0-G800680.  
 XX PR 19-FEB-2000; 2000GB-0004196.  
 XX (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX PA  
 XX Testa TT;  
 XX WP1; 2001-502792/55.  
 DR P-PSDB; AAU07053.  
 PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 PS Claim 5; Page 25-28; 66pp; English.  
 CC The sequence represents a cDNA which encodes a human Flamingo  
 CC polypeptide. Flamingo is a member of the G-protein coupled receptor  
 CC family, which is involved in signal transduction pathways. By screening  
 CC to identify compounds that stimulate or inhibit the function or level of  
 CC the protein, treatments can be developed for various diseases and  
 CC bacterial, fungal, protozoan and viral infections, including HIV, cancer,  
 CC diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease, acute  
 CC heart failure, hypotension, hypertension, urinary retention, angina  
 CC pectoris, myocardial infarction, stroke, ulcers, allergies and benign  
 CC prostatic hypertrophy. Also treatable are psychotic and neurological



CC disorders such as anxiety, schizophrenia, manic depression, delirium,  
CC dementia, severe mental retardation, Huntington's disease and Gilles de  
CC la Tourette's syndrome.

XX Sequence 8871 BP; 1712 A; 2859 C; 2587 G; 1713 T; 0 other:

# Alignment Scores:

Pred. No.:	0	Length:	8871
Score:	2837.00	Matches:	2837
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.06%	Indels:	0
DB:	22	Gaps:	0

US-09-916-849A-3 (1-2923) x AAS11677 (1-8871)

```

QY 1 MetAAGSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
DB 1 ATGCGAGGCCGCGCCACCGGCGTCCCTCCCAACGCGCGCGCTGCTGCTG 60
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAerGlnValGlyProCysAerSerLeu 40
DB 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTTrLeuCysProSer 60
DB 121 GGGTCAGAGGAGCAGAGGCTCTTGCGGGGGCTGCGCCCATGAGGCTGCTGCTCATCC 180
QY 61 SerAlaSerAsnLeuTTrLeuTTrLeuTTrSerArgCysArgAspAlaGlyThrGluLeuThr 80
DB 181 TCAGGCTCAACCTTCGCTCTTACACACACCGCTGAGGAGATGCGGCACTGAGCTGACT 240
QY 81 GlyHisLeuValProHisAspGlyLeuArgValTTrCysProGluSerGluAlaHis 100
DB 241 GGCACCTGCTACCCACCAACATGCGCTGAGGGTTTGGTCTCAGAAATCCAGGCCCAT 300
QY 101 TLeuProLeuProProAlaProGluGlyCysProTTrSerCysArgLeuLeuGlyTLeuGly 120
DB 301 ATTCCCTTACCAACAGCTCTTGAAGGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGluGluHisProCysLeuVala 140
DB 361 GGCACCTTTCACCAAGGAGCAAGCTCACTGCGCGAGAGAGACCCGCTTAAAGGCT 420
QY 141 ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160
DB 421 CCACGCTCAGATGACAGCTCTGCAAGCTGGCAAGGCCCGGGCTGAGGCGAGGAGAA 480
QY 161 ArgSerProGluGluSerLeuGlyValArgGlyValArgAsnValAsnThrAlaProGln 180
DB 481 AGGTCAACCAAGAGATCCCTGGGTGGCGCTCGGAAAGAAATGTAAATACAGCCCCCAG 540
QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200
DB 541 TTCACGCCCCCAAGCTACAGGCGCACAGTCCGGAACCAAGCAAGCAAGCAAGCAAGCTGTT 600
QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGlyTTrMet 220
DB 601 GCATCCCTAAGGGCCATCGACCGCGAGCGAGGGTGAAGGCACTGAGATCAACCAAG 660
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
DB 661 GATGCTCTTGTATATGCGCTTCAACAGATTCCTTCCCTGGAACCAATCTGCTGTGA 720
QY 241 ValTTrThrAlaGluGluLeuAspArgGluTTrLysSerThrHisValPheArgValThr 260
DB 721 GTTACCAACGCGAGAGAGCTGATCTGTGACCAACAGACCAACCTCTTCCAGGGTCAAG 780
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280
DB 781 GCGCAGAGACACGAGATGCGCCGACAGAGTGCCTGGCTACACTCACTCACTTGGTTACT 840
QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTTrLysLeuSerLeuArgGlu 300

```

```

DB 841 GACACCAATGACCAATGACCTGTGTGAGCAGCAGAGTACAGAGAGGCTCAGGAG 900
QY 301 AsnLeuGluValGlyTTrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320
DB 901 AACCTGAGGTGGCTATGAGGTCTCACTGACGGGCAAGATGATGATGATGATGATGATGATG 960
QY 321 AsnAlaAsnIleLeuTTrArgLeuLeuGluGlySerGlyGlySerProSerGluValPhe 340
DB 961 AATGCAATATTTCTGTACCGCTGCTGAGAGGGTCTGGGGGCAACCCCTTGAACTCTT 1020
QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360
DB 1021 GAGATCGACCTCGCTGCGGGGTATCGAACCCTGGGCTCTGTGATCTCGGAAAGAGGTG 1080
QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyAspProGlyProArgSer 380
DB 1081 GATTCCTACAGCTGACCGTGAAGAGCAAGTACCAAGGTGGAGACCCGGGTCTCGAGAT 1140
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPheSer 400
DB 1141 ACCACGCGCGTGTTCCTTCTGTGAGAGATGACATGATTAATGCCCCCAAGTTTATG 1200
QY 401 GlyValArgGlyValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
DB 1201 GAGAAAGCTATGTGTGCTCAGGTGAGGAGATGTGACTCCAGGGGCGCCAGTACTCCGA 1260
QY 421 ValTTrAlaSerAspAspAspGlySerAsnAlaValaHisTyrSerIleMetSer 440
DB 1261 GTACAGCTCTGATCGAGACAGAGAGCAATGCGTGTGCTATATGACTATGATGATGATG 1320
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValaSer 460
DB 1321 GGCATATGCTGGGGAGAGTTTATCTGGAATGCCAGACTGAGAGCTGAGATGTGGTGGC 1380
QY 461 ProLeuAspTyrGluTTrThrLysGluTTrThrLeuArgValaArgAlaGlnAspGlyGly 480
DB 1381 CCTCTTGTACTATGAGCAGCAGCAAGAGTACCCCTGACGGTGCAGCAGAGATGTGTGC 1440
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
DB 1441 CGTCCCCACTCTTATATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
DB 1501 AATGCCCATCTTCTGTGACGACCCCTTCCAGGCTACTGCTGGAAGAGGCTCCCTTA 1560
QY 521 GlyTTrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
DB 1561 GGCTACCTGGTCTTCAGATGTCAGGCTATGACCGCTGATGCTGATGATGATGATGATGATG 1620
QY 541 GluTTrArgLeuAlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyTTrGly 560
DB 1621 GATTAACGCTGCTGCTGGGTGGGATGATGATCTTCCCTTCAACATCAACATGAGCAAGGC 1680
QY 561 TTrIleSerValaAlaGluLeuAspArgGluGluValaAspPheTyrSerPheGlyVal 580
DB 1681 TGGATCTCTGTGGCTGTGAACCTGACCGGAGAGATGATTTCTTACAGCTTGGGGTGA 1740
QY 581 GluIleAspAspHisGlyTTrProAlaLeuThrAlaSerAlaSerValThrVal 600
DB 1741 GAAAGCTCAGACCAAGGAGCTTCAGACTCACTGCTGCGCAAGTCAAGGTGAGCTGCTG 1800
QY 601 LeuAspValaAspAspAsnProThrPheThrGlnProGluTTrValArgLeuAsn 620
DB 1801 CTGATGTCACAGCAACATCAACCTTTCACCAACAGATACAGAGTGGGCTCAT 1860
QY 621 GluAspAlaAlaValGlyThrSerValaThrValSerAlaValAspArgAspAlaHis 640
DB 1861 GAGATGAGCTGTGGGAGCAGAGGTGTGACGGTGTGACGGTGTGACGGTGTGACGGTGTGAT 1920
QY 641 SerValIleTTrTyrGlnIleThrSerGlyValAsnThrArgAsnArgPheSerIleThrSer 660

```



QY 1381 ArggIyLeuArgGlnaArgPheH1sPheThrLeuAlaLeuSerPheAlaThrLySgluArg 1400  
DB 4141 CCGGCGCTCGCGCAGGAGGTTTCCACTTCACTCGCTGCGCTCTCGTTGGCCAAAGAGGCGC 4200  
QY 1401 AapGlyLeuLeuLeuLeuTyraSngLyArGPhaSngLuvH1saSPheValAlaLeuGlu 1420  
DB 4201 GACGGGTGCTGTGTGACATGGCGCTTCAATGAAAGCATGACTTGTGTGCGCTCGAG 4260  
QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLysThrThrThrValSer 1440  
DB 4261 GTGATCCAGAGAGGTTCCAGCTCACTTCTCTGCGAGGGAGTCAACCAACCGGTGTCC 4320  
QY 1441 ProPheValProGlyLyValSerAapGlyGlnTrpH1sThrValGlnLeuLySgLyTr 1460  
DB 4321 CATTCGTGCTCGCGAGAGTCAAGTATGGCCAGTGGCATACGGTGGCAATGAAATATCTAC 4380  
QY 1461 AeuLySProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnLySgLyAla 1480  
DB 4381 AATTAAGCCACTGTGTGGTACAGAGGGCTCCCAAGGGGCCCATCAGAGAGAGAGGTGGCT 4440  
QY 1481 ValValThrValAaSPGlyCySaSPThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
DB 4441 GTGGTACCGGTGGATGGCTGTGTACACAGAGTGGCTTGGCTTGGATCTGTCTCGGCGC 4500  
QY 1501 AeuTySerCyAlaAaLaglnGlyThrGlnGlyGlySerLySgLySerLeuAapLeuThr 1520  
DB 4501 AACTACTCTGTGTGCTCCAGAGGCCACCGAGGTGGCAGAGAAAGTCTGTGATCTGACG 4560  
QY 1521 GlyProLeuLeuLeuGlyGlyLyValProAapLeuProGlyLysThrProValAaPheLeuArg 1540  
DB 4561 GGAGCCCTGTACTAGCGCGGGGCTGTGACTGCGCGAGAGCTTCCAGATCCGAATGCGG 4620  
QY 1541 GlnPheValGlyCySweArGhaLeuGlnValAaSPerArgH1sIleAaPheValAaP 1560  
DB 4621 CAGTTCGTGGGCTGCATGGCGAACTCGAGGTGGAGACCGCGGCACTAGACATGGCTGAC 4680  
QY 1561 PheIleAlaAaSPaSngLyThrValProGlyCySProAlaLySgLyAaSPaValCySaSPer 1580  
DB 4681 TTCAATGGCCAAACATGGCACCGGTGCTGGCTGCGCCAGAGAAAGTGTGTGACAGC 4740  
QY 1581 AeuThrCyH1saSngLyGlyThrCySaValaSngLInTrpAaP1aPheSerCySgLyCys 1600  
DB 4741 AACACTGTGCACAAATGGGGCACTTGTGTAACATGGGAGCGCTTCAAGCTGCAGAGTGC 4800  
QY 1601 ProLeuGlyPheGlyGlyLySerCySaLaglnGlnLeuAlaAaSProGlnH1sPheLeu 1620  
DB 4801 CCCGTGGGCTTTGGGGGCAAGAGCTGCGCCAGAGAAATGGCCAAATCAACAGCACTTCCG 4860  
QY 1621 GlySerSerLeuValAlaTrpH1sGlyLeuSerLeuProIleSerGlnProTrpTyLeu 1640  
DB 4861 GGCAGCAGCCTGGGTGGCTGGCATGGCTCTGCTGCTGCCATCTCCCAACCCCTGTGACCTC 4920  
QY 1641 SerLeuMetPheArGhThrArgGlnAlaAaSPGlyValLeuLeuGlnAlaIleThrArgGly 1660  
DB 4921 AGCTATATGTCCGACGCGCCAGCGCGAGGTGTCTGTGAGGCCATCAACAGGGGG 4980  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlyLyH1saValMetLeuSerValGlyGlyThr 1680  
DB 4981 CGAGAGCAATCACTTACAGCTACAGAGGGGCACTGATGTGAGCCTGGAGGGCAACA 5040  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyValGlnAlaAaSPGlyAaPTrpH1s 1700  
DB 5041 GGGCTTCAGGGCCCTCTCTCTCGTGGAGCCAGCGGGGCCAATACCGTGTGACTGGCAC 5100  
QY 1701 H1saLaglnLeuAlaLeuGlyAlaSerGlyLyProGlyH1sAlaIleLeuSerPheAaP 1720  
DB 5101 CATGCACAGCTGGCATGGAGGCCAGCGGGGGCTGGCCATGCCATCTGTCTTCCAT 5160  
QY 1721 TyTGlyGlnGlnArgAlaGlyLyAaSPLeuGlyProArgLeuH1sGlyLeuH1sLeuSer 1740  
DB 5161 TATGGCAGAGAGAGAGAGGGGCAACTGTGGGCCCGGCTGTGATGTGTGACACTGAGC 5220  
QY 1741 AeuIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaAaSPGlyPheArgGly 1760

DB 5221 AACATAAGTGGCGGAAATACGTGGGCAAGCCGGGTGGGCCCGGTGTGGGGC 5280  
QY 1761 CysLeuGlnGlyValaArgValSerAaPThrProGlyGlyValaAaSPerLeuAaSPerSer 1780  
DB 5281 TGTGTGAGGGGTGGCGGGTGGAGGATACCGCAAGGGGGTTAAACAGCTGGAATCCAGC 5340  
QY 1781 H1sGlyLysSerIleAaSPaLaglnGlyCySaSPerLeuProAaPProCySaSPerAaP 1800  
DB 5341 CATGGGAGAGACATCAACGTGGACCAAGGCTGTAGCTGTGACCTGTGTGACTCAAC 5400  
QY 1801 ProCySProAlaAaSPerTyCySerAaSPaPTrpAaSPerTySerCySerCySaSP 1820  
DB 5401 CCGTGTCTGTCAAGCTATGTGACAGCAACGACTGGGACAGCTTTCCTCGACGCTGTAT 5460  
QY 1821 ProGlyTyTrpGlyAaSPaSPeThrAaSPValCySaAaPLeuAaSPProCySgLyH1sGln 1840  
DB 5461 CAGGGTATCTATGGTGAACCTGTAATGTGTGTGACCTGAACCGGTGAGACACAG 5520  
QY 1841 SerValCyThrArgLySProSerAlaProH1sGlyTyThrCySgLyCySProProAaP 1860  
DB 5521 TCTGTGTATCCCGCAAGCCAGTGGCCCGCATGGCTATACCTGGAGTGTCCCCAAAT 5580  
QY 1861 TyTrLeuGlyProTyCySgLyThrArgGlyLeaSPGlnProCySProArgGlyTrpTrpGly 1880  
DB 5581 TACCTGGGCGATCTGTGAGACAGAGATGACAGGCTTGTCCCGGTGGCTGGGGGA 5640  
QY 1881 H1sProThrCySgLyProCySaSPCySaAaPValSerLySgLyPheAaPProAaSPCySaSP 1900  
DB 5641 CATCCCAATGTGGCCCATGACCTGATGTGATGTGACAAAGGCTTGAACCAAGACTGACAC 5700  
QY 1901 LysThrSerGlyGlyCySwh1sCySgLySgLyAaSPH1sTyTrArgProProGlyLysProThr 1920  
DB 5701 AAGACAGCGCGAGTGGCATGTGACAGAGAGAACCACTACCGGCCCAAGGAGCCCCAAC 5760  
QY 1921 CysLeuLeuCySaAPCySgLyTrpProThrGlySerLeuSerArgValCySaAPProGlyAaP 1940  
DB 5761 TGCCCTGTGTGACTGTATCCCAAGGCTCTTGTGCCAAGTCTGTGACCTGAGAGAT 5820  
QY 1941 GlyGlnCySPProCySgLySProGlyValIleGlyValArgGlnCySaAPArgCySaAPaSPPro 1960  
DB 5821 GGCAGTGTCAATGCAGACAGGTGTCAATCGGGGTCAAGTGTACCGCTGTGACACCT 5880  
QY 1961 PheAlaGlyValaIleThrAaSPGlyCySgLyValaAaSPTrpAaSPerCySPArgAlaIle 1980  
DB 5881 TTGTGTAGTTCACCAACATGGCTGTGTAAGTGAATATGACAGTGGCCACAGCGCAT 5940  
QY 1981 GlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCySPPro 2000  
DB 5941 GAGGTGGGATCTGTGTGGCCCGTAACCGCTTGGGCTGCTGTGCTGCCCTGTCTCC 6000  
QY 2001 LysGlySerPheGlyThrAlaValaArgH1sCySaAPGlyH1sArgGlyTyTrpLeuProPro 2020  
DB 6001 AAGGCTTTTGGGAGCTGTGTGCGCCACTGTATGAGACAGAGGGGTGGCTCCCCCA 6060  
QY 2021 AeuLeuPheAaSPeThrSerIleThrPheSerGlnLeuLySgLyPheAlaGlyArgLeu 2040  
DB 6061 AACCTTCACTGACAGTGCATCACTTCTCAAACTGAAGGGCTTGGCTGTGAGCGGCTA 6120  
QY 2041 GlnArgAaSPGlySgLyLeuAaSPSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg 2060  
DB 6121 CAGCGGAATGAGTCAAGGCTTGAAGGAGGGGCTCCAGAGCTAGAGGCTGTCTCTGTGCG 6180  
QY 2061 AeuAlaThrGlnH1sThrAlaGlyTyTrPheGlySerAaPValLySgLyAlaTyTrGlnLeu 2080  
DB 6181 AACGCCAGCAGACACAGAGCTGTACTTGGCAGCAGACGTCAAGTGGCTTACCAAGT 6240  
QY 2081 AlaThrArgLeuLeuAlaH1sGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
DB 6241 GCCACGCGGTGTGGCCCAAGAGAGCACCCAGGGGGCTTGTGGGCTGTGTGCCACAG 6300  
QY 2101 AaPValH1sPheThrGlnAaSPLeuArgValGlySerAlaLeuLeuAaSPThrAlaAaP 2120

D 6301 GACGTGACCTTCATGAGATCTGCGGGGCGTGGAGGCGCCCTCTGAGACAGCCAC 6360  
O 2121 LysaRghIstrpGluLeuIleIeInGlnThrgIuGIyGlyThrAlaItrpLeuGlnIhI 2140  
D 6361 AAGCGGCACTGGAGCTGATCCAGACAGAGGGTGGACCGCTGGCTGCTGCAGAC 6420  
O 2141 TrrGuaIaIYrAlaSerAlaLeuAlaGlnAaMecIArgHietThrTyrlLeuSerProthe 2160  
D 6421 TATGAGGCTTACGCGAGTCCCTGGCCAGAACATGGCGCACACTTACCTTAAGCCCCCTTC 6480  
O 2161 ThrIleValIthrProasnIleValIleSerValIalArgLeuAapIyGlyAaMpheAla 2180  
D 6481 ACCATCGTCAAGGCCCAACATTGTCACTCCGTAAGTCCCTTGAGAAAGGAACCTTCTCT 6540  
O 2181 GilyAlaIyLeuProArgTrrGluAlaLeuArgGlyGluInProProAapLeuGluThr 2200  
D 6541 GGGGCGAAGCTGCCCCGCTACAGAGCCCTGCGTGGGAGAGCCGCCGAGACCTTGAGACA 6600  
O 2201 ThrValIleLeuProGluSerValIpheArgGluThrProProValIalArgProAlaGly 2220  
D 6601 AAGTCACTTCTGCTGAGTCTGTCTTCAGAGAGAGCCGCCGAGGTGAGGCCGCGAGGC 6660  
O 2221 ProGlyGluAlaGlnIuProGluGluLeuAlaArgArgGlnArgArgHietProGluLeu 2240  
D 6661 CCGGAGAGGCGCCAGAGCCAGAGAGGCTGGCACCGGCAAGCAGCAGCCGAGAGCTG 6720  
O 2241 SerGlnGlyGluAlaValAlaSerValIleIeTyArgThrLeuAlaGlyLeuLeuPro 2260  
D 6721 AGCCAGGGTGAAGGCTGGTGGCCAGCTCATCTACCCGACCCCTGGCGGCTACTGCT 6780  
O 2261 HisAntYrAapProAapIyAArgSerIeLeuArgValIProIyAArgProIleIleAsnThr 2280  
D 6781 CATACATrAGACCCCTGACAGAGGAGCTTGAGAGTCCCAAGCCCATCATACACACA 6840  
O 2281 ProValIalSerIleSerValIhisAapAapGluGluLeuLeuProArgAlaLeuAapIy 2300  
D 6841 CCGGAGTGAACATCAGGCTCATATATGAGAGGCTTCTGCCCCGGGCGCTTGACAA 6900  
O 2301 ProValIthrValIalGlnPheArgLeuLeuGluThrGluGluArgThrIyProIleCyValI 2320  
D 6901 CCGCTCACGGTGCAGTTCCTGGCTGTCAGACAGAGGAGCGGACCAAGCCCATCTGATCTC 6960  
O 2321 PheTrpAaMhisSerIleLeuValIserGlyThrgIyGlyTrpSerAlaArgGlyCyGlu 2340  
D 6961 TTTCTGGAACATTCATATCTGATCTGATGAGTGCACAGGTGCTGTGGCCAGAGCTGAGAA 7020  
O 2341 ValValIpheArgAaGlnIuSerHietAlaSerCyGlnIuCyAaMhisAmetThSerPheAla 2360  
D 7021 GTGCTCTTCGCAATGAGAGGCAAGTCAAGTCCAGTCCACACATGACAGAGCTTCGCT 7080  
O 2361 ValIleuMecAapValIserArgArgGluAaGlnIyGluIleLeuProLeuIyThrLeuThr 2380  
D 7081 GTGCTCATGAGCGTTCCTCGCGGAGATGGGAGATCCCTGCCATGAGACACTGACA 7140  
O 2381 TyrValIalLeuGlyValIThrLeuAlaIleuLeuLeuThrPhePheLeuThrLeu 2400  
D 7141 TACGTGGCTTACGAGTGTCACTTGGCTGCTCTTCTGCTCACCTTCTTCTCACTCTC 7200  
O 2401 LeuArgIleLeuAaGSerAaGlnIhietGlyIleArgArgAaMecIThrAlaIleuGly 2420  
D 7201 TTGGGTATCTCGCCCTCCACACACAGCGCATCCGACTAACCTGACAGCTCCCTGGGC 7260  
O 2421 LeuAlaGlnLeuValIpheLeuLeuGlyIleAaGlnIalAaPLeuProPheAlaCyThr 2440  
D 7261 CTGCTCAAGCTGAGCTTCTCTCTGAGATCAACCAAGGCTGACCTCCCTTGTGCTGACA 7320  
O 2441 ValIleAlaIleLeuLeuHisApheLeuTyrlLeuCyThrPheSerTrpAlaLeuLeuGlu 2460  
D 7321 GTCACTTGCATCTGCTGACATCTCTGACCTTGCACCTTCTCTGAGCTGCTGAG 7380  
O 2461 AlaLeuHisAleuTyArgAlaLeuThrgIuValIArgAapValIaMThrgIyProMecArg 2480  
D 7381 GCCTTGACCTGTACCGGGGCACTCACTGAGGTGGCGATGTCAACACCGGCGCCCATGCGC 7440

O 2481 PheTyrlYrMecLeuGlyTrpGlyValIProAlaPheIleThrgIyLeuAlaValGlyLeu 2500  
D 7441 TTTCACTACATGCTGGGCTGGGGGCTGCTGCTTCAATCAAGGGCTrAGCGGTGGCTG 7500  
O 2501 AapProGluGlyTyrrGlyAaMProAapPheCyrrPheLeuSerIleTyrrAapThrLeuIle 2520  
D 7501 GACCCGAGGGCTACAGGAACCTTGACTTGTGGCTCTTCATCTTATGACACGCTCATTC 7560  
O 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMecSerValIpheLeuTyrlIleu 2540  
D 7561 TGGAGTTTGTGGCCCGGTGGCCCTTTGGCTTCGATGAGTGTCTTCTTATCATCTG 7620  
O 2541 AlaAlaArgAlaSerCyValAlaGlnArgGlnIyPheGluIyGlyGlyProValSer 2560  
D 7621 GCGGCGCGGCTCTCTGTGCTGCCAGCGGAGGGCTTTGAGAGAAAGGCTGTCTCG 7680  
O 2561 GilyLeuGlnProSerPheAlaValaLeuLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
D 7681 GGCTTGAGGCTTCTTCTGCGCTCTCTCTCTGCTGAGCGCCAGTGGCTGTGCACTG 7740  
O 2581 LeuSerValAaMserAapThrLeuLeuPheHietTyrlLeuPheAlaThrCyAaMcyAlle 2600  
D 7741 CTCTCTTCAACAGCCACACCTCTCTTCACTCACTCTTGTGCTTACTGCAATTGATC 7800  
O 2601 GlnGlyProPheIlePheLeuSerTyrrValIleuSerIyGlyValaArgIyAlaLeu 2620  
D 7801 CAGGCGCCCTTCACTTCTCTCTCTATGTGTGTGACAGAGAGGTCCGGAAGCACTC 7860  
O 2621 LyIleuAlaCySerAaGlyIyProSerProAapProAlaLeuThrThrIySerThrLeu 2640  
D 7861 AAGCTTGTGAGCGCGCAAGCCCAAGCCCTGACCTGTGACCAACCAAGTCCACCTG 7920  
O 2641 ThrSerSerTyrrAaMcyProSerProTyrrAlaAapGlyArgLeuTyrrGlnProTyrrGly 2660  
D 7921 ACCCTGCTTCACTACCTCCCAAGCCCTTACGAGATGGGGCTGTACAGACCTTACGGA 7980  
O 2661 AapSerAlaGlySerLeuHisAserThrSerArgSerGlyIyLysSerGlnProSerTyrrIle 2680  
D 7981 GACTCGGCGGCTCTCTGACACACACAGTGTGCTGGGCAAGAGTCAAGCCAGCTATCATC 8040  
O 2681 ProPheLeuLeuArgGluGluGluSerAlaLeuAaMProGlyGlnGlyProProGlyIyLeuGly 2700  
D 8041 CCTCTTGTGTGAGAGAGAGTCCGACATGACCTGAGCAAGGGCCCTTGGCTGTGGG 8100  
O 2701 AapProGlySerLeuPheLeuGluGlyGlnAaPglnglnIhietAaPProAapThrAapSer 2720  
D 8101 GATCCAGGAGGCTGTCTTCTGGAAGGTCAAGACACAGACATATCTTGACAGGACTCC 8160  
O 2721 AapSerAapLeuSerLeuGluAaPAPglnglnIySerTyrrAlaSerThrHisSerSer 2740  
D 8161 GACAGTACCTGTCTTAGAAGAGACAGAGTGGCTCTATGCTCTTCACTCACTCATCA 8220  
O 2741 AapSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
D 8221 GACAGTGAAG 8280  
O 2761 TrpAapSerLeuLeuGlyProGlyAlaGluArgLeuProLeuHisSerThrProIyAap 2780  
D 8281 TTGGATAGCTTGTGGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 8340  
O 2781 GilyGlyProGlyIyProGlyIyValaProTrpProGlyIyAapPheGlyThrThrAlaIyGly 2800  
D 8341 GGGGGCCAGGGGCTGTGCAAGGCGCCCTGGCCAGGAGACTTTGGGACACAGCAAAAGAG 8400  
O 2801 SerSerGlyAaMglValaProGluGluArgLeuArgGluAaMglValaAaPLeuAaM 2820  
D 8401 AGTAGTGAACAGGGGCGCTGAGAGAGCGCTGCGAGAGATGAGATGCGCTGTCTCGA 8460  
O 2821 GlnGlySerLeuGlyProLeuProGlyIySerSerAlaGlnProHisIyGly 2837  
D 8461 GAGGGGTCCTTAGGCCCTTCCAGGCTTTCGCGCCAGCTCAACAAAGGT 8511





619 uAenGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAl 639  
DB CAATAGAGATGACGCTGGGGGACACGCGTGTGACGGGTGACGCTGGACCGATGAC 818  
QY aHisSerValIleThrTyrrGlnIleThrSerGlyValenThrArgAspArgPheSerIleGln 659  
DB TCATATGTCATCACTTACCAATCAACAGTGGCAATCTCAAAACCGGTTCTTCATCAAC 878  
QY rSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGlnLys 679  
DB CAGCCAAAGTGGTGGGCTGGTATCCCTTCCTGCACTGACCTCAAACTTGAGCG 938  
QY 619 GglnTyrrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
DB GCAGATATGTTGGCTGTTACCGCTCCGATGGCACTCGGACAGACACGGACCAATGTT 998  
QY 1ValAlaenValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
DB GGTGATATGACCGACCGCCACCAACCTGCTCTTCAAGAGCTCCCACTATACAGT 1058  
QY 719 1AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGly 739  
DB GAATGTTATAGAGACCGGCGCGACGACCAAGGTGGTGGTGGTGGTGGTGGTGGTGGT 1118  
QY 739 uAspThrGlyGluAsnAlaArgIleThrTyrrPheMetGluAspSerIleProGlnPheArg 759  
DB GGACCAAGGTGAGATGCGCCGATCACCTTCAATGAGAGACACATCCCGCACTTCG 1178  
QY 759 GtleAspAlaAspThrArgValAlaValThrThrGlnAlaGlnLeuAspTyrGluAspGlnVal 779  
DB CATCATACAGACACGGGGGCGTGTCAACACCGGCTGAGCTGAGCTGAGCTGAGCTGAG 1238  
QY 779 1SerTyrrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrTh 799  
DB GTCTTACACCTGGGCTGCTTACTGCTCGGGACATATGGCATTCCTCCAGAAAGTCCGACACAC 1298  
QY 799 rTyrrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgPhe 819  
DB CTACCTGAGATCCCTGGTGAACGAGTGAATGACATGCCCCCTCAGTTCCTGGCAGACCTC 1358  
QY 819 rTyrrGlnIleSerValTyrrGluAspValProProPheThrSerValIleGlnIleSerAla 839  
DB CTACCAAGGAGAGTGTCTATGAGAGATGTGCACCTTCACTAGCGCTCGACGATTCCTGCG 1418  
QY 839 aThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrrThrPheGlnIleGlyLysAspAs 859  
DB CACTCATGTGATTTCTGACCTTAAAGGAGGCTCTTCAACCTTCCAGAGGCGGACGACGA 1478  
QY 859 rGlyAspGlyAspPheIleValGlnSerThrSerGlyIleValAlaArgThrLeuAspArgLe 879  
DB TGGAGACGGTGACTTATGTTGATGATCCAGCTCAGGACATCGGACGATACGGAGGCT 1538  
QY 879 uAspArgGluAsnValAlaGlnTyrrValLeuArgAlaTyrrAlaValAspLysGlyMetPr 899  
DB GGAATGAGAGAACGCGGCCAGTATGTCTTGGGGCATATGACAGTGAACAGGGGATGCC 1598  
QY 899 oProAlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProPr 919  
DB CCGAGCCCGGACACCTATGAGAGTGAAGTCACTGTGTGATGATGATGATGATGATGATGAT 1658  
QY 919 oValPheGlnGluAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAla 939  
DB TGTCTTTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718  
QY 939 aValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrrGly 959  
DB CGTGGCCCGGGGTCACAGCCACGACCCGATGAGAGACCAATGCGCAGATTATGATACCA 1778  
QY 959 nIleValGlnGluAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
DB GATGTGTGAGGAGCAACATCCCTGAGGTCTTCAAGTGAACATCTTCTCCGGGAGAGCTGAC 1838  
QY 979 rAlaLeuValAspLeuAspTyrrGluAspArgProGluTyrrValLeuValIleGlnAlaThr 999

1839 AGCCCTGGTACCTTACCTACGAGAGACCGGCTGAGTACGCTCGTCACTCCAGGAC 1898  
QY rSerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAs 1019  
DB GTCACTCTCTGTGTGAGCCGCGGCTACAGTCAAGTCCGCTCTTCAAGCCGAAAGACAA 1958  
QY 1019 nProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrrValThrAsnArgSerSe 1039  
DB CCGACAGGTGTGGGCAACTTATGATTCCTTTCAACACTATGTCACCAATCCGTCACAG 2018  
QY rSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSe 1059  
DB CAGCTTCCCTGGGGGTGTCATGGCCAGATGACGCCCATACCTGATATCTCAGATAG 2078  
QY 1059 rLeuThrTyrrSerPheGlnArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
DB TCTGACTTACAGCTTTTACGCGGGGAAATGAATCAAGCTGTGGTCTTCAATGCTTCAC 2138  
QY rGlyGluLeuLysSerArgAlaLeuAspAsnAspArgProLeuGlnAlaIleMetSe 1099  
DB GGATGAGCTGAGCTTACCGCGGACCTGACACAAACCGGCTCTGAGAGGCAATCAAG 2198  
QY 1099 rValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuAspArgValThrI 1119  
DB CGTGTGGGTGACAGCGCGGTGACACAGCTGACCGCCAGTGGCGGCTGGGTGACCAT 2258  
QY 1119 eIleThrAspGlyMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGly 1139  
DB CATCAAGATGAGATGTCACCCACACATCAACCTGCGCTGGAGAGACATGACCCGA 2318  
QY 1139 uArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaThrLeuAlaThr 1159  
DB GCGCTTCTGTGACCACTGCTAGGCTCTTCAATCAAGCGGTGGCCCGCACGCTGGCCAC 2378  
QY 1159 rProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
DB GCCACCGGACCAAGTGTGTGCTTCAAGTGAAGGAGACACCGACCCCGGGGGGACCA 2438  
QY 1179 eIleLeuAsnValSerLeuSerValGlnIleProProGlyIleGlyIleProProPh 1199  
DB CATCTCAAGCTGAGCTGTGTGGGCGAGCGCCGACGGGCTCGGGGCGGCGCTT 2498  
QY 1199 eLeuProSerGlyAspLeuGlnGluArgLeuTyrrLeuAsnArgSerLeuThrAlaI 1219  
DB CTTGCCCTTGAAGACCTGAGAGAGGCCCTTACTCAACCGAGCCTGCTGAGAGGCCAT 2558  
QY 1219 eSerAlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCysGluAs 1239  
DB CTGGGACACAGCGCGTGTGCCCTTCCAGACCAACATCTGCTGGGGAGGCTCGAGAA 2618  
QY 1239 nTyrrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleHisSerSe 1259  
DB CTACATGCGCTGCGCTGTGGGCTGCTTCACTCTCGCGGCTTCACTGCTCTCTC 2678  
QY 1259 rSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
DB CTCCGTGCTCTTCCGGGCCATCCACCCCTCGAGAGGCTGCGCTGCGGCTGCGGCGG 2738  
QY 1279 yPheThrGlyAspTyrrCysGlnThrGluValAspLeuCysTyrrSerArgProCysGlyPr 1299  
DB CTTCAAGGGTGAAGTACGACGACCGAGGTGACCTCTGCTCACTCGGCGGCTGTGGCC 2798  
QY 1299 oHisGlyArgCysArgSerArgGlnGlyGlyTyrrThrCysLeuCysArgArgAspGlyTyrrTh 1319  
DB CCACGGGCGGCTGCGGACCGGCGGAGGGGCTTACCTGCTGTGTGATGAGCTTAC 2858  
QY 1319 rGlyGlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysLysAs 1339  
DB GGATGACACATGTGAGGTGATGCTGCTCAAGCGCTGAGACCCCGGTGTGTGACAGAA 2918  
QY 1339 nGlyGlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAs 1359



Db 2919 TGGGGGACCTGTGTCAACCTGTGGTGGCGGTTTCAAGTGCATTCGCCATCTGGAGA 2978  
 QY 1359 pPheGlyuysProGlyrCyseGlnValThrThrArgSerPheProAlaHisSerPhe1Leh 1379  
 Db 2979 CTTCGAGAAAGCCCTATGCGCAGGTGACACGCGCGCTTCCCGGCACCTCTTCATCAAC 3038  
 QY 1379 rPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLeuG1 1399  
 Db 3039 CTTTGGCGGCGCTGGCGCAAGCTTTCACCTTCACCTCGGCCCTCTCGTTGGCCCAAGGA 3098  
 QY 1399 uArgAspGlyLeuLeuLeuLeuTyrrAsnGlyArgPheAsnGlyLysPheValAlaLe 1419  
 Db 3099 GGGCGACGCGGTGTGCTGTGTACATGGCGCTTCAATGAGAGCATGATGTTGTGGCCCT 3158  
 QY 1419 uGlnVal11IleGlnGlnVal1GlnLeuThrPheSerAlaGlyGlnSerThrThrVal 1439  
 Db 3159 CAGAGTATTCAGAGACAGAGTCCAGCTCACTCTCTGCAAGGAGGTCAACCAACGAGT 3218  
 QY 1439 lSerProPheValProGlyGlyValSerAspGlyGlnTrpHisThrValGlnLeuLysTy 1459  
 Db 3219 GTCCCACTTCGTGGCCGAGAGTCACTATGGCCAGTGGCATACGTTGCACTGAATA 3278  
 QY 1459 rTyrrAsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnLysVa 1479  
 Db 3279 CTACAAATAGCCACTGTGTGGTCAAGACAGGCTCCCAAGGCGCCATCAGACAGAGAGGT 3338  
 QY 1479 lAlaValValThrValAspGlyCyseAspThrGlyValAlaLeuArgPheGlySerValLe 1499  
 Db 3339 GGGTGTGTGATCCGTGGATGGCTGTGTGACACAGAGTGGCGCTTGGCATCTGTCT 3398  
 QY 1499 uGlyAsnTyrrSerCyseAlaAla-glnGlyThrGlnGlyLysSerLysSerLeuAspL 1519  
 Db 3399 GGGCAACTACTCTGTGTGTGCCCAAGGACCAAGGTTGGCAGCAAGATCTCTGAATC 3458  
 QY 1519 eutThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlnSerPheProValArgM 1539  
 Db 3459 TACAGGAGGCGCCCTGTCTACTAGAGGAGGAGTCTGTGACTGCCCAAGACTTCCAGTCCGA 3518  
 QY 1539 eArgGlnPheValGlyCyseMetArgAsnLeuGlnValAspSerArgHis1LeaMetA 1559  
 Db 3519 TGGGAGATTCGTGGGCTGCATTCGGAACTTCAGGTGACACCGGCACATAGACATGG 3578  
 QY 1559 lAspPhe1LeaAlaAsnAsnGlyThrValProGlyCyseProAlaLysLysAsnValCyseA 1579  
 Db 3579 CTGACTTCACTTGCACAAATGGCACCGTGCCTGGCTGCCCAAGAGAAAGTGTGTG 3638  
 QY 1579 gSerSerAsn-ThrCyseHisAsnGlyGlyThrCyseValAsnGlnTrpAspAlaPheSerCyse 1598  
 Db 3639 ACAGCAA-AACTTGCACAAATGGGAGCACTTGGGTAAACCAAGTGGACCGCTTCAGCTGC 3697  
 QY 1599 GlnCyseProLeuGlyPheGlyGlyLysSerCyseAlaGlnGlnMetAlaAsnProGlnHis 1618  
 Db 3698 GAGTGGCCCTTGGGCTTGGGGGCAAGACTGGGCCCAAGAAATGCCAATCCACAGCAC 3757  
 QY 1619 PheLeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuPro1LeSerGlnProTrp 1638  
 Db 3758 TTCCTGGGACGACGCTGTGTGGCTGCATGGCTCTCGCTGCCATCTCCCAACCTGTG 3817  
 QY 1639 TyrrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAla1Leh 1658  
 Db 3818 TACTCAAGCTCAATGTTCCGACGCGCCACAGCGAGGTCTGTGACAGGCATCACCC 3877  
 QY 1659 ArgGlyArgSerThr1LehThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGln 1678  
 Db 3878 AAGGGCGGACGACCACTACCTTACAGCTTACAGAGGGCCAGTGTATGTGACGCGTGGAG 3937  
 QY 1679 GlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAsp 1698  
 Db 3938 GGGACAGGGCTTACAGGCTCTCTCTCGCTGAGACACAGCGGGCCCAATGACGCTGAC 3997  
 QY 1699 TrpHisHisAlaGlnLeuAlaLeuGlyAlaSer-glyGlyProGlyHisAla1LeuSer 1718  
 Db 3998 TGGCACTACGACAGCTGGACCTGGAGGACAT-CGGGGGGCTTGGGCAATGCCATCTGTTC 4056

QY 1718 rPheAspTyrrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHis 1738  
 Db 4057 CTTCATATATGGGACGAGAGAGAGAGGCAACTGGGCGCCCGCGCTGCATGGCTTGC 4116  
 QY 1738 sLeuSerAsn1LehThrValGlyGly1LeProGlyProAlaGlyGlyValAlaArgGlyPh 1758  
 Db 4117 CCGTGAACAACATTAAGTGGCGGAAATACCTGGGCGAGCGGCGGTGTGGCCGCTT 4176  
 QY 1758 eArgGlyCyseLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAs 1778  
 Db 4177 TCGGGGCTGTGTGACGGGTGTGGCGGTGAGCGAATACGCAAGGAGGCTTAAACGCTGGA 4236  
 QY 1778 pProSerHisGlyGlnSer1LeaAsnValGlnGlnGlyCyseSerLeuProAspProCyseAs 1798  
 Db 4237 TCCAGCCATGGGAGAGACATCACTGAGACAGAGCTGTAGCTGTGACCTTGTG 4296  
 QY 1798 pSerAsnProCyseProAlaAsnSerTyrrCyseSerAsnAspTrpAspSerTyrrSerCyse 1818  
 Db 4297 CTCAACCCCGTGTCTGTCTAACACTATTCAGACAGACTGGGACAGCTATTCCTGACG 4356  
 QY 1818 rCyseAspProGlyTyrrTyrrGlyAspAsnCyseThrAsnValCyseAspLeuAsnProCyseG1 1838  
 Db 4357 CTGTGATCCAGGTTACTATGTGTGACAACTGTACTAATGTGTGACCTGAACCCGTGGA 4416  
 QY 1838 uHisGlnSerValCyseThrArgLysProSerAlaProHisGlyTyrrThrCyseGlnCysePr 1858  
 Db 4417 GCACCACTGTGTGTATCCCGACAGGCCAGTGGCCCCCATATGGCTTATCTGAGGTTC 4476  
 QY 1858 oProAsnTyrrLeuGlyProTyrrCyseGlnThrArg1LeaAspLanProCyseProArgGlyTy 1878  
 Db 4477 CCAAAATTACTTGGGCACTACTGTGAGACCAAGATTTGACAGGCTTGTCCCCGAGTGC 4536  
 QY 1878 pTrpGlyHisProThrCyseGlyProCyseAsnCyseAspValSerLysGlyPheAspProAs 1898  
 Db 4537 GTGGGACATCCACATGTGGCCCAATGCATGTGAATGACAGAAAGCTTTCACCA 4956  
 QY 1898 pCyseAsnLysThrSerGlyGlnCyseHisCyseLysGlnLysAsn1sTyrrArgProProGlySe 1918  
 Db 4597 CTGCAACAAGACAAAGCGCGAGTGCATCTGCAAGAGAACCACTACCGGCCCCACGAG 4656  
 QY 1918 rProThrCyseLeuLeuCyseAspCytyrrProThrGlySerLeuSerArgValCyseAspPr 1938  
 Db 4657 CCCCACCTGCTCTGTGTGTGACTGTACCCCAAGGCTCTTGTCTCAGAGCTGTGACC 4716  
 QY 1938 oGlnAspGlyGlnCyseProCyseLysProGlyVal11IleGlyArgGlnCyseAspArgCyseAs 1958  
 Db 4717 TGAAGATGGCAAGTGTCCATGACAGCGAGGTGTATGGGCGTCAAGTGTGACCGCTGTGA 4776  
 QY 1958 pAsnProPheAlaGlnValThrThrAsnGlyCyseGlnValAsnTyrrAspSerCyseProAr 1978  
 Db 4777 CAACCTTTTCTGAGGTCAACCAACAAATGGCTGTGAAGTGAATTTATGACACTGCCACG 4836  
 QY 1978 gAla11IleGlnAlaGly1LeTrpTrpProArgThrArgPheGlyLeuProAlaAla1Asp 1998  
 Db 4837 AGCAATTAAGCTGGAACTGTGTGGCCCGTAAACCGCTTCCGGGCTGTGTGCTGCC 4896  
 QY 1998 oCyseProLysGlySerPheGlyThrAlaValArgHisCyseAspGlnHisArgGlyTyrrPhe 2018  
 Db 4897 CTGTCCAAAGGCTCTTGGGAACTGTGTGGCGCACTGTATGAGCACAGGGGGTGTCT 4956  
 QY 2018 uProProAsnLeuPheAsnCyseThrSer1LehThrPheSerGlnLeuLysGlyPheAlaG1 2038  
 Db 4957 CCCCCCAACCTTTCAACTGACGTCATCACTTTCAGAACTGAAGGGCTTGCCTGA 5016  
 QY 2038 uArgLeuGlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLe 2058  
 Db 5017 GCGGCTTACAGGAGATGTGTGAGGCTTGAAGTCAAGGGGCTTCCACAGCAGTACGCTGTCT 5076  
 QY 2058 uLeuArgAsnAlaThrGlnHisThrAlaGlyTyrrPheGlySerAspValLysValAlaTy 2078  
 Db 5077 CCGTGGCAACCGCACGACGACACAGCTGGCTATCTTGGGACAGCAAGTGTGGCTTA 5136

2078 rglndleuAlaThrArgleuLeuAlaHisgluSerThrThrArglyPheglYleuSerAl 2098  
 5137 CCAAGTGGCCACAGCGCGCTGGCCACAGAGCACCCAGCGGCGCTTGGGCTGTCTGGC 5196  
 2098 aThrGlnAspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThr 2118  
 5197 CACACAGAGCGTGCATTCACCTGAGAAATCTGCTGGGGTGGGACAGGCCCTTCCTGGACAC 5256  
 2118 rAlaAsnLysArgHisTrpGluLeuLisgluGlnThrGluGlyGlyThrAlaTrpLeuLeu 2138  
 5257 AGCCCAACAGCGCGCTGGAGCTGATCCAGCAGACAGAGGGTGGCACCGCTGGCTGCT 5316  
 2138 uGlnHisTrpGlyuAlaTrpAlaSerAlaLeuAlaGlnAsnMetArgHisTrpTrpLeuSe 2158  
 5317 CCAAGACTATATAGAGCTATAGCGCAGATGCTGGCCCAAGAACATGGCGGCACACTTACTTAA 5376  
 2158 rProPheThrIleValThrProAsnIleValIleSerValValArgLeuAspLysGlyAs 2178  
 5377 CCCCTTCAACATTCGTACAGCCCAACATTTGTATCTCCGTAGTGCCTTGGACAAAGGGA 5436  
 2178 nPheAlaGlyValAlaLysLeuProArgTrpGlyuAlaLeuArglyGlyGluGlnProProAspLe 2198  
 5437 CTTTGCTGGGGCCAAAGCTGCCCGGCTACAGAGCCCTGGGTGGAGAGAGCCCGGCAACT 5496  
 2198 uGluThrThrValIleLeuProGluSerValPheArgGluThrProProValValArgPr 2218  
 5497 TGAACAACAAGTCAATCTGCTGAGTCTGTCTTCAAGAGAAGCGCCCGCTGGTCAAGGCC 5556  
 2218 oAlaGlyProGlyGlyuAlaGlnGluProGluGluAlaArgArgGlnArgArgHisPr 2238  
 5557 CGAGAGCCCGGAGAGGCCCAAGAGCCAGAGAGAGCTGGCAGCGGACAGGCAAGCCAGCC 5616  
 2238 cGluLeuSerGlnGlyuAlaValAlaSerValIleIleTrpArgThrLeuAlaGlyLe 2258  
 5617 GGAACCTGAGCGAGGGTGAAGCTGTGGCCAGGCTCATCTACCGACCCCTGGCCGGGCT 5676  
 2258 uLeuProHisAsnTrpAspProAspLysArgSerLeuArgValProLysArgProIleI 2278  
 5677 ACTCCCTCATATACATAGACCTTGAACAGCGCAGCTTGAAGTCTCCCAAAACCCCGATCAT 5736  
 2278 eAsnThrProValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLe 2298  
 5737 CAACACACCCCTGGTGGAGCATCACGCTCCATGATGATGAGAGCTTCTGCCCCGGGCTCT 5796  
 2298 uAspLysProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrLysProI 2318  
 5797 GGAACAACCCCTGACGCGTGAAGTCCGCTGCTGAGAGCAGAGAGGAGGCAAGGCCCAT 5856  
 2318 eCyValPheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTrpSerAlaArgG 2338  
 5857 CTGTGTCTTCTGGAACCAATTCATCTGTGTCAATGGGCAAGGTGGCTGGTGGCCAGAGG 5916  
 2338 YCyGluValValPheArgAsnGluSerHisValSerCyAsnGlnCyAsnHisMetTrpSe 2358  
 5917 CTGTGAAGTCTCTTCCGCAATGAGGCACTGACCTGCGACAGTGGCAACATGAGAGAG 5976  
 2358 rPheAlaValLeuLeuAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysTrh 2378  
 5977 CTTGCTGTGTCTATGAGCGTTTCTCGGCGGAGAAATGGGAGATCTCGCACTGAAGAGAC 6036  
 2378 rLeuThrTrpValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLe 2398  
 6037 ACTACACATACCTGGCTCTTAAAGTGTACCTTGGCGCCCTTCTGTCTCACCTTCTTCTTCT 6096  
 2398 uThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAl 2418  
 6097 CACTCTCTTGGGTATCTCGGCTCCAAACCAACAGGCACTCGAATCTGACAGCTGC 6156  
 2418 aLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAl 2438  
 6157 CTTGGGGCTGGCTAGCTGTGTCTTCTCTGGAGATCAACAGGCTGACCTCCCTTCTTTCG 6216  
 2438 aCyThrValIleAlaIleLeuLeuHisPheLeuTrpLeuCyThrPheSerTrpAlaLe 2458

6217 CTGCACAGTATGCGCATCTCGCTGACATTCCTGTACTCTGACACTTTCCTGGGCTCT 6276  
 2458 uLeuGluAlaLeuHisIleLeuTrpArgAlaLeuThrGluValArgAspValAsnThrGlyPr 2478  
 6277 GCTGAAGGCCCTTGCACCTGTACCGGGCACTCACTGAAGGTGGCAATGTCAACACCGGCC 6336  
 2478 oMetArgPheTrpTrpMetLeuGlyTrpGlyValProAlaPheIleThrGlyLeuAlaVala 2498  
 6337 CATCGCTTCTACTACATAGCTGGGCTGGGGGTGGTCCGCTTATTCACAGGGCTAGCCGT 6396  
 2498 lGlyLeuAspProGluGlyTrpGlyAsnProAspPheCyTrpLeuSerIleTrpAspThr 2518  
 6397 GGGCTGAGACCCCAAGAGGCTTACAGGAAACCTGACTTGTGGCTGTCTCATCTATGACAC 6456  
 2518 rLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrp 2538  
 6457 GCTCATCTGAAGTGTCTGAGCCGAGTGGCTTGTGGCTTGTCAATGAGTGTCTTCTGTGA 6516  
 2538 rIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGlyuLysLysGlyPr 2558  
 6517 CATCTGGCGGCGCGGCGCTCTCTGTGTGCTGCCAGCGGCAAGGCTTGAAGAAAGTCTCC 6576  
 2558 oValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLe 2578  
 6577 TGTCTCGGGCTGGCAGGCCCTTCTGCGCGTCTCTGCTGCTGAGCCGACAGTGGCTGCT 6636  
 2578 uAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTrpLeuPheAlaThrCyAs 2598  
 6637 GGCATCTGCTCTGTGTCAACAGCAACACTCTCTTCACTACATCTTCTTCTACCTGGCA 6696  
 2598 nCyAlleGlnGlyProPheIlePheLeuSerTrpValValLeuSerLysGlyValArgly 2618  
 6697 TTGCATCAAGGGCCCTTATCTTCTCTCTCTATGTGTGCTTACGAAGAGGTCCGGA 6756  
 2618 sAlaLeuLysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSe 2638  
 6757 AGCATCTCAACCTTGCCTGCGCGGCAAGCCCAAGCCCTGACCTGTCTGACCAACAGTC 6816  
 2638 rThrLeuThrSerSerTrpAsnCyAspProSerProTrpAlaAspGlyArgLeuTrpGlnPr 2658  
 6817 CACCTGACCTCTGCTCTTACATCTGCCCCAGCCCTTACGAGATGGGGCTGTACCAAGCC 6876  
 2658 oTrp--GlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnPro 2677  
 6877 CTA-ACGAGACATCGGCGGCTCTGTGACAGCAAGCAAGTGTCTGGCAAGAGTCAAGCCC 6935  
 2678 SerTrpIleProPheLeuLeuArgGluGluSerAlaLeuAsnProGly 2693  
 6936 AGCTACATCCCTTCTTGTGTGAGGAGAGTCCGCACTGAACCTGGG 6983

RESULT 6  
 ABLK15177  
 ID ABLK15177 standard; DNA; 9321 BP.  
 XX  
 AC ABLK15177;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human REPTR 9 cDNA sequence.  
 XX  
 KW REPTR; human; antiinflammatory; cyrostatic; immunosuppressive;  
 KW antiviral; anti-HIV; antitubercular; anticonvulsant; nootropic;  
 KW neuroprotective; antiallergic; antibody; immunogen; endometritis;  
 KW gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KW Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KW endocrine disorder; hypochalams disorder; Kallman's disease;  
 KW autoimmune disease; inflammatory disease; infertility; receptor;  
 KW acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KW allergy; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KW systemic lupus erythematosus; cell proliferative disorder;  
 KW cancer; developmental disorder; Duchenne muscular dystrophy; gene;  
 KW Becker muscular dystrophy; neurological disorder; epilepsy;

[illegible][illegible]

Oy 393 AsnAspAsnAlaProGlnPheSerGluValArgGlyValValGlnValArgGluAspVal 412  
 Db 1402 AATGATAATGCCCCCAGTTTATGTAGAGAGCGCTATGTGTCTCAGATGAGGAGATGTG 1461  
 Oy 413 ThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAla 432  
 Db 1462 ACTCCAGAGGGCCCCAGTACTCCGAGTCACAGCTCGATCGAGACAGAGGAGGACATGCTC 1521  
 Oy 433 ValValHisGlySerTyrLeuSerGlyValAsnAlaArgGlyGlnPheTyrLeuAspAlaGln 452  
 Db 1522 GTGGTGCACTATAGCATCATGATGATGCGCAATGCTCGGGGACAGTTTATCTGATGCCAG 1581  
 Oy 453 ThrGlyAlaLeuAspValValSerProLeuAspGlyGlyGlnThrThrGlyGluTyrThrLeu 472  
 Db 1582 ACTGAGGCTCTGAGAGTGTGTGAGCCCTCTTGACTATGAGACAGACAGAGAGTACACCTTA 1641  
 Oy 473 ArgValAlaGlyAlaGlyAspGlyGlyValArgProLeuSerAsnValSerGlyValLeuValThr 492  
 Db 1642 CCGGTGCGAGACAGAGATGTGTGCGCTGCCCATCTCTATATGTCTGTGGCTTGAGTGA 1701  
 Oy 493 ValGlnValLeuAspGlyLeuAspAsnAlaProIlePheValSerThrProPheGlnAla 512  
 Db 1702 GTACAGGCTCTGAGATATCAAGACAAATGCCCATCTTGTCAGACACCCCTTCCAGGCT 1761  
 Oy 513 ThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAla 532  
 Db 1762 ACTGTCCGAGAGAGCGTCCCTTAGGCTACCTGTCTTCATGTCTCAGGCTATCCAGCCT 1821  
 Oy 533 AspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPhePro 552  
 Db 1822 GATGCTGTGACATATGCCCGCTGAGATACCGCTTGCTGGGGTGGAGACATGACTTCCC 1881  
 Oy 553 PheThrIleAsnAsnGlyTyrGlyTyrIleSerValAlaAlaGluLeuAspArgGlyGlu 572  
 Db 1882 TTCACATCAACAATGAGCAGAGGCTGAGTCTGTGGCTGTGAGCTGAGCCGGAGGAA 1941  
 Oy 573 ValAspPheTyrSerPheGlyValGlnValArgAspHisGlyThrProAlaLeuThrAla 592  
 Db 1942 GTGATTTCTACAGCTTGGGGGTAGAACCTCGACGACATGGACCTCCAGCACTCACTCC 2001  
 Oy 593 SerAlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGln 612  
 Db 2002 TCCGCGCATGTCAAGGTACTCTCTGATGTCAACGACACATCAACTTATCCAA 2061  
 Oy 613 ProGluTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrVal 632  
 Db 2062 CCAAGGTACACAGTCCGCTCAATGAGATGAGCTGTGGGACAGAGTGTGACGCTG 2121  
 Oy 633 SerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThr 652  
 Db 2122 TCAGCTGTGACCCGTGATGCTCATGTGTCTACCTTACCCAGATCACAAGTGGCAATACT 2181  
 Oy 653 ArgAsnArgPheSerIleThrSerGlnSerGlyGlyValLeuValSerLeuAlaLeuPro 672  
 Db 2182 CGAAACCCCTTCTCATCAACAGCCAAAGTGTGTGGCTGTGATCCCTTGCCCTGCA 2241  
 Oy 673 LeuAspTyrTyrLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArg 692  
 Db 2242 CTGAGCTACAAACTTGAGCGGAGATGTGTGTGGCTGTATCCGCTCGAGTGGCACTCG 2301  
 Oy 693 GlnAspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPhe 712  
 Db 2302 CAGAGCAGGACCAATGTGTGTGTATGTACCCGACGCAACCCATCGTCTGTCTTT 2361  
 Oy 713 GlnSerSerHisGlyTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValVal 732  
 Db 2362 CAGAGCTCCCATATACAGTGAATTTATATGAGACCGGCGGACGAGCAACAGTGTGTG 2421  
 Oy 733 LeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGlu 752  
 Db 2422 CTGATCAAGCGCACGATGAGGACACAGGTGGAATGCCGATCACTTACATGAG 2481  
 Oy 753 AspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlu 772

Db 2482 GACAGCATCCCCCAGTTCGCGATCGATGACAGACAGGGGCTGTACCAACCAAGCTGAG 2541  
 Oy 773 LeuAspTyrGlyAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIle 792  
 Db 2542 CTGAGCTACGAAAGACCAAGTCTTATCAACCTGTGCCTATTACTGTCTGGGAAATGGCAAT 2601  
 Oy 793 ProGluTyrSerAspThrThrTyrLeuGluIleLeuValAsnAspValAspAsnAla 812  
 Db 2602 CCCAGAAAGTCCGACACCACTTACCTGAGATCTTGAGTGAACGATGAAATGACATGCC 2661  
 Oy 813 ProGlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProPhePheThr 832  
 Db 2662 CCTCAGTCTCCGACAGCTCCTTACCAAGGCGAGTGTCTATGAGATGTGCCACCTTCACT 2721  
 Oy 833 SerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyr 852  
 Db 2722 AGCTTCTGAGATCTCAAGCCATGATGTGATTTCTGACTTAATGGCGAGGCTTCTTAC 2781  
 Oy 853 ThrPheGlnGlyIleAspAspGlyAspGlyAspPheIleValGlnSerThrSerGlyIle 872  
 Db 2782 ACTTCCAGAGAGCGACGATGAGAGCGGTACTTATTTGTGTGATCCACTCAGGATC 2841  
 Oy 873 ValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyr 892  
 Db 2842 GTGCGAAGCTTACGAGAGCTGATCGAGAGAACGTGGCCGATGTCTTGGCGGCAATAT 2901  
 Oy 893 AlaValAspIleGlyMetProProAlaArgThrPheMetGluValThrValThrValLeu 912  
 Db 2902 GCACTGGAACAAGGAGAGTCCCCCAGCCCGCACCTATGAAAGTACATCATGTGTG 2961  
 Oy 913 AspValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGlu 932  
 Db 2962 GATGTGATGACATATCCCTCGTCTTTGAGCAGAGTGAATTTATGTGTGTGGAAGAG 3021  
 Oy 933 AsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThr 952  
 Db 3022 AACAGCCCAATGGGCTGAGCGGTGCCCGGCTCACACCTGACCCCGATGAAGGACCC 3081  
 Oy 953 AsnAlaGlnIleMetTyrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAsp 972  
 Db 3082 AATGCCAGATTTATGTACAGATTTGTGAGAGGCAACATCTCTGAGTCTTCCACTGGAC 3141  
 Oy 973 IlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGlyAspArgProGluTyr 992  
 Db 3142 ATCTTCCGGGGAGCTGACAGCCCTGTGTGACTTATCACTACAGAGACCGGCTGAGTAC 3201  
 Oy 993 ValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArg 1012  
 Db 3202 GTCTGTGATCCAGGCGACAGTCAAGCTCTCTGTGTAGCCGGGCTTACAGTCCAGTCCGC 3261  
 Oy 1013 LeuLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAsn 1032  
 Db 3262 CTCTTTCACCGCAATGACAAACCAACAGTGTGGGCAACTTGTGATCTTTTCAACAC 3321  
 Oy 1033 TyrValThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHis 1052  
 Db 3322 TATGTCAACATGCTCAACAGAGTCTCCGTGGGGTCCATTGGCCGAGTACCTGCCCAT 3381  
 Oy 1053 AspProAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeu 1072  
 Db 3382 GACCTATATCTCAGATGATGTGACTTACAGCTTTTGAAGCGGGGAATGAATCAAGCTCG 3441  
 Oy 1073 ValLeuLeuAsnAlaSerThrGlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArg 1092  
 Db 3442 GTCTTCTCAATGCTTCAAGGATGAGCTTAAGCTTAAGCCGCGCATGAGACACAAACCG 3501  
 Oy 1093 ProLeuGluAlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGln 1112  
 Db 3502 CCTGTGAGGCGCATCATGACGATGTGTGTGTGACAGCGGCTGACACAGCGTGAACGCCAG 3561  
 Oy 1113 CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArg 1132

Db 3562 TGGCGCTGCGTGTGACCATCATCAACCGATGAGTGTCTCAACCAAGCATCACGCTGCGC 3621  
 Qy 1133 LeuGluAspMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAla 1152  
 Db 3622 CTGGAGGACATGTACACCGAGCGCTTCTGTGCACACCTGCTAGAGCTCTTCATATCCAGCGC 3681  
 Qy 1153 ValAlaAlaThrLeuAlaThrProProAspHisValAlaValPheAsnValGlnArgAsp 1172  
 Db 3682 GTGGCGCGCACGCTGGCGCACCGCACCGGACCAAGTGTGTCTTCACAGTCACAGCGGAGC 3741  
 Qy 1173 ThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGly 1192  
 Db 3742 ACCGACGCCCGCGGGGCGCATCTCTCAACGTGAGCTGTCTGGGGCGCGCGCCAGCGG 3801  
 Qy 1193 ProGlyGlyLeuProPheLeuProSerGlyLeuAspLeuGlnGlyLeuGlyLeuAsn 1212  
 Db 3802 CCGGGGGGGGGCGCGCTTCTGTCTCTGTAGGACCTGACGAGGCGCCCTATATCTCAAC 3861  
 Qy 1213 ArgSerLeuLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCys 1232  
 Db 3862 CGCAGCTGTGCACGCGCATCTCGGACAGGGGGTGTGCTTCACAGCAACATCTGC 3921  
 Qy 1233 LeuArgGluProCysGlyAsnIleTyrMetArgCysValSerValLeuArgPheAspSer 1252  
 Db 3922 CTGCGGGAGCGCTGCGAGAACTACATGCGCTGTCTGTCTGGCTGCTCTGCTCTCC 3981  
 Qy 1253 AlaProPheIleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeu 1272  
 Db 3982 GCGCGCTTATCGCT 4041  
 Qy 1273 ArgCysArgCysProProGlyPheThrGlyAspTyrCysGlnThrGluValAspLeuCys 1292  
 Db 4042 CGCTGCGCGTGCCTGCGCGCTTCAAGGGTGACTACTGCGAGACCGAGGTGGACCTTGC 4101  
 Qy 1293 TyrSerArgProCysGlyProHisGlyArgCysArgSerArgGlyGlyGlyTyrThrCys 1312  
 Db 4102 TACTCGGGCGCTGTGTGCGCCCAAGGGCGCTGCGCGCGCGCGCGCTCTACCTGC 4161  
 Qy 1313 LeuCysArgAspGlyTyrThrGlyGlnHisCysGlyValSerAlaArgSerGlyArgCys 1332  
 Db 4162 CTCTGTCTATATGCTCTACCGGGTGAACCTGTAGAGTGTAGTCTCTCTCTCTCTCTCT 4221  
 Qy 1333 ThrProGlyValCysIleAsnGlyGlyTyrCysValAsnLeuValGlyGlyPheIle 1352  
 Db 4222 ACCCGGGTGTGTGCAAGAAATGGGGGCACTGTGTCAACCTGCTGGGCGGTTTCAAG 4281  
 Qy 1353 CysAspCysProSerGlyAspPheGlyLeuPheCysTyrCysGlnValThrThrArgSerPhe 1372  
 Db 4282 TGCATTTGCCATCTGTGAGACTTGTGAGAACCTTACCTGCAAGGTGACCAAGCGCGACTTC 4341  
 Qy 1373 ProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAla 1392  
 Db 4342 CCGCGCCACTCTTCATCACTTTCGCGGCTGTGCGCAAGGCTTTCACCTTCACTTGC 4401  
 Qy 1393 LeuSerPheAlaThrIleGlyArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGly 1412  
 Db 4402 CTCTCGTTTGTCCAAAGAGAGCGGAGCGGCTGTCTGTGTACAAATGGGCGTTTCAATGAG 4461  
 Qy 1413 LysHisAspPheValAlaLeuGlnValIleGlnGlnGlnValGlnLeuThrPheSerAla 1432  
 Db 4462 AAGCATATCTTGTGGCGCTGTGAGGTATTCAGAGAGCGGTCCAGCTTCTCTGCA 4521  
 Qy 1433 GlyLeuSerThrThrThrValSerProPheValProGlyGlyValSerAspGlyGlnTyr 1452  
 Db 4522 GGGGAGTCAACCAACCAAGGTGTCCCATTTGGTGTCCCGAGAGAGTCAATGATGGCAATGG 4581  
 Qy 1453 HisThrValGlnLeuGlyTyrTyrAsnIleProLeuLeuGlyGlnThrGlyLeuProGln 1472  
 Db 4582 CATACGATGACGCTGAATATCTACAAATAGCCACTGTGTGGTCAAGAGGGCTCCCAAG 4641  
 Qy 1473 GlyProSerGlnGlyValAlaValAlaValThrValAspGlyCysAspThrGlyValAla 1492  
 Db 4642 GGGCCATCAAGACAGAAAGTGTGTGTGTATACCTGTGTATGCTGTGTACAGAGAGTGGC 4701

Qy 1493 LeuArgPheGlySerValIleuGlyAsnIleTyrSerCysAlaAlaGlnGlyThrGlnGlyGly 1512  
 Db 4702 TTGGCTGTGATCTGTCTGTGGCACTACTCTGTGTGTGCGCCAGGGCAACCCAGGGTGGC 4761  
 Qy 1513 SerIleYsSerLeuAspLeuThrGlyProLeuLeuLeuGlyGlyValProAspLeuPro 1532  
 Db 4762 AGCAAGAACTCTGTGATCTGACGGGGCCCTGTCTACTAGGGGGGTGTGCTGACCTGCC 4821  
 Qy 1533 GluSerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAsp 1552  
 Db 4822 GAGAGCTTCCAGTCCGAATGGCGGCGAGTTCGTGGGTGTGATGGGAACCTGCAAGTGGAGC 4881  
 Qy 1553 SerArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyTyrValProGlyCysPro 1572  
 Db 4882 AGCCGACATATGACATAGCTGTACTTATTCACCAATGGCAACCGTGTGCTGCTCCCT 4941  
 Qy 1573 AlaIleYsAsnValCysAspSerAsnIleCysHisAsnGlyGlyTyrCysValAsnGln 1592  
 Db 4942 GCCAAGAAAGACGTGTGTGACAGCAACCTTGCCACAAATGGGGGCACTTGCGTGAACAG 5001  
 Qy 1593 ThrAspAlaPheSerCysGlyCysProLeuGlyPheGlyGlyIleYsSerCysAlaGlnGly 1612  
 Db 5002 TGGAGCGCTTCAAGCTGACAGATGCCCCCTTGGGGGCGAAGAGCTGCGCCAGAGAA 5061  
 Qy 1613 MetAlaAsnProGlnHisPheLeuGlySerSerLeuValAlaTyrHisGlyLeuSerLeu 1632  
 Db 5062 ATGGCAATCCACAGACATCTTCTGGGACAGACCTGTGTGTGCTGTGCTCTGCTG 5121  
 Qy 1633 ProIleSerGlnProTyrPheLeuSerLeuMetPheArgThrArgGlnAlaAspGlyVal 1652  
 Db 5122 CCCATCTCCCAACCTGTGATCTCAAGCTCATGTTCGCGAGCGGCGCAAGCGGTGTC 5181  
 Qy 1653 LeuLeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHis 1672  
 Db 5182 CTGTGCAAGCCATACCAAGGGGCGAGGCGAGCAACATCACTTCAAGTACAGAGGGCCAC 5241  
 Qy 1673 ValMetLeuSerValGlnGlyTyrGlyLeuGlnAlaSerSerLeuArgLeuGlyProGly 1692  
 Db 5242 GTGATCTGATAGCTGTGAGGGGCAAGGGCTTCAAGGCTCTCTCTCTCTCTCTCTCT 5301  
 Qy 1693 ArgAlaAsnAspGlyAspTyrHisHisAlaGlnLeuAlaLeuGlyValAspGlyIlePro 1712  
 Db 5302 CGGGCCAAATGACGCTGTACCTGACCAATGACAGCTGTGAGCCAGCGGGGGCTT 5361  
 Qy 1713 GlyHisAlaIleLeuSerPheAspTyrGlnGlnArgAlaGlnGlyValAsnLeuGlyPro 1732  
 Db 5362 GGCATGCAATCTGTCTCTTCAATTAATGGGCAAGAGAGCAAGAGGGCAACCTGGGCCCC 5421  
 Qy 5422 CCGCTGCATAGCTGTGACCTGAGCAACATTAACAGTGGGGGAATATCTGGGCGAGCGGG 5481  
 Qy 1753 GlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlu 1772  
 Db 5482 GGTGTGGCCCGTGTGTGTGGGTGTGTGCAAGGGGTGTGTGCGGTGTGCAATACGCCAGAG 5541  
 Qy 1773 GlyValAsnSerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSer 1792  
 Db 5542 GGGGTTTAAAGCTGTGATCTCCAGCCATGGGGAGAGCATCAAGTGAAGCAAGGCTTAC 5601  
 Qy 1793 LeuProAspProCysAspSerAspProCysProAlaAsnSerTyrCysSerAsnAspTyr 1812  
 Db 5602 CTGCTGACCTTGTGATCAAAACCGGTCTCTGTACAGCTAATGAGCAACGACTGG 5661  
 Qy 1813 AspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAspAsnGlySerHisValCys 1832  
 Db 5662 GACAGCTATCTCTGACGCTGTATTCAGGTACTATAGTGAACAACGTACTAATGTGTGT 5721  
 Qy 1833 AspLeuAsnProCysGlyGlnGlnSerValCysThrArgIleAspProSerAlaProHisGly 1852  
 Db 5722 GACTTAACCTCGTGTAGACACCAAGTGTGTGTGTATACCGGACAGCCAGTCCCCCAGTGG 5781



OY	1853	YrYrthrCysGluCysProProbaSerIleuGlyProTyrCysGluThrArgIleAspGln	1872
Db	5782	TATACCTCCAGAGTGTCCCCAAATTACCTTGGCGCATCTGTGAGACCAAGATTGACAG	5841
OY	1873	ProCysProAspGluTyrTrpGlyHisProThrCysGlyProCysAsnCysAspValSer	1892
Db	5842	CCATTGTCCCCGTGGGTGGAGATCCCAATGTGGCCCAATGCATCTGAATGTCAAG	5901
OY	1893	LysGlyPheAspProAspCysAsnLysThrSerGlyGluCysHisCysGlyLysAsnHis	1912
Db	5902	AAAGGCTTTGACCCAGACTGCAGACAAAGCAAGCGCGAGGTGCCACTGCAAGAGAACAC	5961
OY	1913	TyrArgProProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeu	1932
Db	5962	TACCGGCCCCAGAGGACGCCCACTGGCTCTGTGTGACTGTACCCCAAGGCTCTTGG	6021
OY	1933	SerArgValCysAspProGluAspGlyGlnCysProCysIleProGlyValIleGlyArg	1952
Db	6022	TCCAGAGTCTGTGACCTCGAGGATGGCCAGAGTGTCCATGCAGACCAAGGTGTATGGGCGT	6081
OY	1953	GlnCysAspArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGlu-----	1970
Db	6082	CAGGTGACCCGCTGTGACAAACCTTTTCTGAGGTCAACCAACATGGCTGTGAAGGGGCC	6141
OY	1971	--ValAsnTyrAspSer--CysProArgAlaIle-Glu--AlaGlyIleTrpTrpProAr	1988
Db	6142	TTGTTTG-CT-AG-TT-ACTGTCCCCG-GCCCA-TGAGGTCTGGCC-TC-CA-GC-AG	6190
OY	1988	GThr-----Arg---PheIly-----LeuProAla-AlaAlaProCysPr	2000
Db	6191	AACCTCTCAGCCAGCTCAGGGGGCTCTCT-GTGTGTCTCCCTGA-GGGCGG-CCCTT-TT	6246
OY	2000	OlyAspIlySerPhe-----GlyThrAlaValArgHisCysAspGlnHisArgGlyTyrPleu	2018
Db	6247	-GG-CT-TCTCTCCCCCGAGGATCTGTCTGGCCACTGTATGAGACAGGGGGTGGCTC	6303
OY	2019	ProProbaLeuPheAsnCysThrSerIleThrPheSerGlyLeuGlyPheAlaGlu	2038
Db	6304	CCCCCAAACTCTTCAACTGCACGTCCTCACTTCTCAGAACTGAAAGGGCTTGGCTGAG	6363
OY	2039	ArgIleuGlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuLeu	2058
Db	6364	CGGCTACAGCGGAATGATCAGGGCTTAACCTCAGGGGGCTCCCAAGCTAAGCCCTCTCC	6423
OY	2059	LeuArgAsnAlaTrnGlnHisIleThrAlaGlyTyrPheGlySerAspValLysValAlaTyr	2078
Db	6424	CTGGCGCAACGCGACGACGACGACGACGACTGGACTTCTGGCAGCGACGTCAAGGTGGCTAC	6483
OY	2079	GlnLeuAlaThrArgLeuLeuAlaHisGlnSerThrGlnArgGlyPheGlyLeuSerAla	2098
Db	6484	CAGCTGGCAAGCGGGCTGTGGCCCAAGAGACCCAGCGGGGCTTGGGCTGTCTGGC	6543
OY	2099	ThrGlnAspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThr	2118
Db	6544	ACACAGGACGCGCACTTCACTGAAATCTGTGGGGTGGGCAAGCGCCCTCTGGACACA	6603
OY	2119	AlaAsnLysArgHisIleTrpGlnLeuIleGlnGlnThrGlnGlyGlyIleThrAlaTrpLeuLeu	2138
Db	6604	GCCAAACAAGCGGCACTGGAGCTTATCCAGGACAGAGGGGTGGACCGCTGGCTGTCTC	6663
OY	2139	GlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisTrnTyrLeuSer	2158
Db	6664	CAGACACTATGAGGGCTAAGCCAGTGGCCCTGGCCAGAAACATGGGCAACTTACTTAAAC	6723
OY	2159	ProPheThrIleValThrProAsnIleValIleSerValValaArgLeuAspLysGlyAsn	2178
Db	6724	CCCTTTCACATCGTCACGCGCCCAATTTGTCTCTCCGTAAGTGCCTTTGGACAAAGGAAAC	6783
OY	2179	PheAlaGlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlyGlnGlnProAspLeu	2198
Db	6784	TTTGTCTGGGGCAAGCTGGCCCCGCTGACAGGCGCTGTGTGGGAGGACGCCCCGGAACTTT	6843
OY	2199	GlnThrThrValIleLeuProGlnSerValPheArgGlnThrProProValValaArgPro	2218

Dh	6844	GAGACAAACGTCATTCTTCGCTCGAGCTCTGTTCAAGAGAGAGGCCCTCCGTCGTGACGCC	6903
Qy	2219	AlaGlyProGlyGluAlaGlnGluProGlyGluLeuAlaArgArgGlnArgArgHisPro	2238
Dh	6904	GCAAGCCCCGGAGAGGCCCAAGAACCAAGAGAGCTGGCAACGCCACACGCGACGCCACCCG	6963
Qy	2239	GluLeuSerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeu	2258
Dh	6964	GAGCTGACCAAGGGGTGAAGGCTGTGGCCAGCTCTCATCTCAACCCGACCTCGGCGGGCTA	7023
Qy	2259	LeuProHisAaenTyrAspProAspGlyArgSerLeuArgValProLysArgProIle	2278
Dh	7024	CTGCTCTAATCACTATGACCTCGACAAAGCCAGCTTGAGAGTCCCAACGCCCATCATC	7083
Qy	2279	AenThrProValValSerIleSerValHisAspAspGlyGluLeuLeuProArgAlaLeu	2298
Dh	7084	AACACACCCCGTGAAGCATCAGCGTCCATGATGATGAGAGGCTTCGCCCGGGCCCTG	7143
Qy	2299	AspLysProValThrValGlnPheArgLeuGlnThrGlyGluArgThrLysProIle	2318
Dh	7144	GACAAACCCGTCACGGGTGACGATTCGCCCTGTGTGAGACAGAGAGGCGAGCCAAACCCATC	7203
Qy	2319	CysValPheTyrPheMetHisSerIleLeuValSerGlyThrGlyGlyTyrPserAlaArgGly	2338
Dh	7204	TGTGTCTTCTGGAAACATTCATCTCTGTCACTGGCACAGTGGCTGTGTGGCCAGAGGC	7263
Qy	2339	CysGlyValValPheArgAenGlySerHisValSerCysGlnCysAaenHisMetThrSer	2358
Dh	7264	TGTGAAGTCGCTTCCGCAATGAGAGCAAGTCAGCTGACAGTGCACACATGACAGAGC	7323
Qy	2359	PheAlaValLeuMetCAspValSerArgArgGlu-----	2369
Dh	7324	TTCCTGTGCTCAATGACGTTTCTTCGCGGAGAGTGGGCCCAAGGCGACGCTGCAGAG	7383
Qy	2370	-----AenGlyGluIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThr	2387
Dh	7384	CCGTGAAATGGGAGATCTGCGCACTGAACAACCTGACATACAGTGGCTCTAGGTCATCC	7443
Qy	2388	LeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsn	2407
Dh	7444	TTCGTGCTCTTCTGCTCACTCTTCTCTCTCACTCTCTGGATCTCGCGCTCCAC	7503
Qy	2408	GlnHisGlyIleArgArgAenLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu	2427
Dh	7504	CAACACGGCAATCCGACGTAACTGACAGCTGCCCTGGGCTGAGCTGAGCTTCTCTC	7563
Qy	2428	LeuGlyIleAenGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis	2447
Dh	7564	CTGGGAATCAACCAAGGTGACCTCCCTTGTGCTGCACAGTCATTTGCCATCTGCTGCAC	7623
Qy	2448	PheLeuTyrLeuCysThrPheSerThrPalaLeuLeuGlnAlaLeuHisIleLeuTyrArgAla	2467
Dh	7624	TTCCTGTAACTCTGACACTTTTCTGTGGGCTCTGTGAGGCCCTTGACCTGTAAACGGGCA	7683
Qy	2468	LeuThrGlyValArgAspValAenThrGlyProMetArgPheTyrTyrMetLeuGlyTyrP	2487
Dh	7684	CTCACTGAGGTGCGCGATGTCAACACCGGCCCCCATGCGCTTCACTACATGACTGGGCTGG	7743
Qy	2488	GlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGlyGlyTyrGlyAaen	2507
Dh	7744	GAGCTGCTGCTCTCAACACAGGCTAGCCGTGGGCTGGAGCCCAAGGGCTACAGGGAAC	7803
Qy	2508	ProAspPheCysTyrLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProVal	2527
Dh	7804	CCTGACTTCTGCTGGCTCTCCATCTATGACACGCTCATCTGAGATTTCGTGGCCCGGAG	7863
Qy	2528	AlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla	2547
Dh	7864	GCTTTTCGCTTCGATGAAGATGTCTTCGTGTAACTCTGGCGGCGCGGCTCTCTGTGCT	7923
Qy	2548	AlaGlnArgGlnGlyPheGlyLysGlyProValSerGlyLeuGlnProSerPheAla	2567



Db 7924 GCCCAGCGGAGGCTTTGAGAAAGATCTGCTCGGGCTTGAGCCCTTCCTCGCC 7983  
 Qy 2568 ValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValaAsnSerAspThr 2587  
 Db 7984 GTCTCTCTGCTGAGCGCCAGCGGCTGCGCTGCTCTGTCGACACGACGCC 8043  
 Qy 2588 LeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
 Db 8044 CTCCTCTTCACCTACCTTTGCTACCTGCAATTGATCCAGGGCCCTTCATCTTCCTC 8103  
 Qy 2608 SerTyrValValLeuSerGlyValArgValAlaLeuLeuAlaCysSerArgVal 2627  
 Db 8104 TCTTATGTGTGTGTAGCAAGAGTCCGAAAGCACTCAAGCTTCTGACGCCCAAG 8163  
 Qy 2628 ProSerProAspProAlaLeuThrTrpLeuSerThrLeuThrSerTyrAsnCysPro 2647  
 Db 8164 CCAGCCCTGACCTGCTGACACCACTCCACCTGACCTGCTGCTCAACCTGCC 8223  
 Qy 2648 SerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHis 2667  
 Db 8224 AGCCCTCAGCAGATGGGGCGGTGTACAGCCCTACGGAGACTCGGCGGCTCTGAC 8283  
 Qy 2668 SerThrSerArgSerGlyLysSerGlnProSerTyrTrlProPheLeuLeuArgGln 2687  
 Db 8284 AGCACAAGTGTGTGCGGCAAGAGTCCAGCCAGCTACATCCCTTCTGCTGAGGAGAG 8343  
 Qy 2688 SerAlaLeuAspProGlyGlnGlyProProGlyLeuGlnYAspProGlySerLeuPheLeu 2707  
 Db 8344 TCCGCACTGACCTGAGCCCAAGGCCCCCTGCGCTGGGGATCCAGGCAAGCTGTTCCTG 8403  
 Qy 2708 GlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspSerLeuGln 2727  
 Db 8404 GAAGGTCAAGACCCAGCAGCATGATCCGACAGCATCCGACAGTGCCTTATGAA 8463  
 Qy 2728 AspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlyGlnGlnGln 2747  
 Db 8464 GACGACCAAGAGTGTCTCTATGTCTCTCACTCATCAAGTACAGAGGAGGAGGAG 8523  
 Qy 2748 GlnGlnGlnGlnGlnAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGlyPro 2767  
 Db 8524 GAGGAGGAGGAGGAGGCGCCCTTCCTGAGAGCAAGGCGTGGATGCTGCTGGGGCT 8583  
 Qy 2768 GlyAlaGlnArgLeuProLeuHisSerThrProLysAspGlyGlyProGlyProGlyLys 2787  
 Db 8584 GAGGAGAGAGGAGCTGCGCCCTGCAAGTATCCCAAGATGGGGGCCCAAGGCTGCAAG 8643  
 Qy 2788 AlaProTrpProGlyAspPheGlyThrThrAlaLysGlnSerSerGlyAsnGlyAlaPro 2807  
 Db 8644 GCCCCTGGCCAGAGACTTTGGGACCAAGCAAAAGAGTGTGTGCAACCGGGCCCT 8703  
 Qy 2808 GlnGlnArgLeuArgGlnLysAspAlaLeuSerArgGlnGlySerLeuGlyProLeu 2827  
 Db 8704 GAGGAGCGGCTGCGGAGAGATGAGATGCCCTCTCTCGAGAGGGGCTCCCTAGGCCCTT 8763  
 Qy 2828 ProGlySerSerAlaGlnProHisLysGlyTyrLeuLysLysCysLeuProThrIle 2847  
 Db 8764 CCAAGGCTCTTCTGCTCCAGCTCAAAAGGATCTTAAAGAGAGTGTGCTGCCACATC 8823  
 Qy 2848 SerGlnLysSerSerLeuLeuArgLeuProLeuGlnGlyCysThrLysSerSerArgGly 2867  
 Db 8824 AGGAGAGAGAGGAGCTCTGCGGCTCCCTCTGAGAGATGCAAGAGGCTTCTCCGAGGC 8883  
 Qy 2868 SerSerAlaSerGlnGlySerArgGlyGlyProProProArgProProArgGlnSer 2887  
 Db 8884 TCTCTCGCTGTGTAGGGCAGCCGGGGGCGCCCTCTCCGCCCAAGCGCCGCGAGAG 8943  
 Qy 2888 LeuGlnGlnGlnLeuAsnGlyValMetProIleAlaMetSerIleLysAlaGlyThrVal 2907  
 Db 8944 CTCACAGAGAGCTGAACGGGGTCATGCTCCATGCGATGACATCAAGAGGAGCAGGCTG 9003  
 Qy 2908 AspGlnAspSerSerGlySerGln 2915  
 Db 9004 GATGAGGACTGCTCAGGCTCCGA 9027

RESULT 7  
 ID AACT6401  
 XX AACT6401 standard; cDNA, 9121 BP.  
 AC AACT6401;  
 XX  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF1956 polynucleotide sequence SEQ ID NO:3911.  
 XX  
 XX Human; open reading frame; ORF; detection; cytosolic; hepatotropic;  
 XX vulnery; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 XX anticonvulsant; osteoplastic; antidiabetic; immunosuppressant; cardiac;  
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 XX hypotensive; dermatological; immunosuppressive; antineoplastic;  
 XX antiviral; antibacterial; antifungal; antineuritic; antihypertensive;  
 XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 XX cholesterol ester storage; systemic lupus erythematosus; infection;  
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 XX allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 XX thrombosis; contraceptive; ss.  
 OS Homo sapiens.  
 XX  
 XX  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX  
 PF 31-MAR-2000; 2000MO-US08621.  
 XX  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CUDAGEN CORP.  
 XX  
 PI Shinketsu RA, Leach M;  
 XX  
 XX WPI; 2000-602362/57.  
 DR P-PSDB; AAB42192.  
 XX  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX  
 PS Claim 5; Page 3061-3067; 5507pp; English.  
 XX  
 CC AACT6446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnery;  
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteoplastic; anticonvulsant; antidiabetic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineuritic;  
 CC antihypertensive; antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 9121 BP; 1797 A; 2892 C; 2529 G; 1901 T; 2 other, SQ

### Alignment Scores:

Pred. No.:	0	Length:	9121
Score:	1623.00	Matches:	1622
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	55.53%	Indels:	0
DB:	.21	Gaps:	0

US-09-916-849A-3 (1-2923) x AAC76401 (1-9121)

QY	1301	G1YARGCYSAASGSRATGAGLUG1YGL1YTYThrCYsleuCYSAAGAAAG1YTYThrG1Y	1320
Db	2527	GGCGGCTGCCACACCCGAGAGGGCGGCTACACCTGCTCTGTGTAATGCTTACACGGGT	2586
QY	1321	GLuHICysGsluVal1SerAlaArgSerG1YARGCYThrProG1YValCYsLYsAsnG1Y	1340
Db	2587	GAGCACTGTGAGGTGAGAGTCTCGCTCAGAGCGCTTGCACCCGGGGTGTCTGCAGAAATGGG	2646
QY	1341	G1YThrCYsValAlaAsnIleuIleuValG1YG1YPhelYsCYsAspCYsProSerG1YAspPhe	1360
Db	2647	GGACCTGTGTCAACCTGCTGTGTGGGTTCAGATGCGAATTCGCCATCTGGAGACTTC	2706
QY	1361	G1YLYsProTYrCYsGlnVal1ThrThrArgSerPheProAlaHisSerPhe1LeThPhe	1380
Db	2707	GAGAAGCCCTAATGCTCCAGGTGACACACGGACGACTTCCCGGCCACTCTCTTATCACTTT	2766
QY	1381	ArgG1YLeuArg1uArg1uArgPheHisPheThrLeuAlaLeuSerPheAlaThrLYsG1YArg	1400
Db	2767	CGGGCCCTGGCCACGCGTTTCCACTTACCTTCGCGCTCTCTGTGGCCACAAAGAACGC	2826
QY	1401	AspG1YLeuIleuLeuTYrTAAsnG1YArgPheAsnG1YLYsHisAspPheVal1AlaLeuG1Y	1420
Db	2827	GACGGGTTCCTGTGTATCAATGGCGCTTTCATATAGAAACATGACTTTGTGGCCCTGAG	2886
QY	1421	Val1LeGlnG1uGlnVal1GlnLeuThrPheSerAlaG1YLeuSerThrThrThrVal1Ser	1440
Db	2887	GTGATCCAGAGAGCAGGTTCACAGCTTCACTTCTCTCAAGGAGGTCAACACACAGGTGTCC	2946
QY	1441	ProPheVal1ProG1YG1YVal1SerAspG1YG1YTrpHisThrVal1G1YIleuLYsTYrTYr	1460
Db	2947	CCATTCTGTCCCGGAGAGATCAGTGAAGGCCACAGTGCATACAGAGCTGAATAACTTAC	3006
QY	1461	AsnLYsProLeuLeuG1YGlnThrG1YLeuProGlnG1YProSerG1uGlnLYsVal1Ala	1480
Db	3007	AATATGACCATGTTGGGTCAAGACAGGGCTCCACAGGGCCCATATAGACAGAAAGTGGCT	3066
QY	1481	ValVal1ThrVal1AspG1YCYsAspThrG1YVal1AlaLeuArgPheG1YSerVal1LeuG1Y	1500
Db	3067	GTGTGTGACCGGTGATGCTGTGACACAGAGATGGCTTGGCTTCGATCTGTCTCGGGC	3126
QY	1501	AsnTYrSerCYsAlaAlaGlnG1YThrG1uG1YG1YSerLYsLYsSerLeuAspLeuThr	1520
Db	3127	AACATCTCTGTGTCTGCCACAGGACACCCAGGGTGTGACAGAAAGTCTCTGAACTTACG	3186
QY	1521	G1YProLeuLeuLeuG1YG1YVal1ProAspLeuProG1uSerPheProVal1ArgMetArg	1540
Db	3187	GGGGCCCTGTACTAAGCGGGGTGCTGACCTGCGCCGAGAGCTTCCCATGTCGGAATGGG	3246
QY	1541	G1uPheVal1G1YCYsMetArgAsnLeuGlnVal1AspSerArgHis1LeaPMeVal1Asp	1560
Db	3247	CAGTTCGTGGCTGCACAGCGGAACCTGCAAGGTGACAGCGGCACATATGACATGTGGCTGAC	3306
QY	1561	Phe1IleAlaAsnAsnG1YThrVal1ProG1YCYsProAlaLYsLYsAsnVal1CYsAspSer	1580
Db	3307	TTCTATTCACCAATATGGCACCGTGTCTGTGGCTGCCCTGCGCAAGAAAGAACTGTGTGACAGC	3366
QY	1581	AsnThrCYsHisAlaAsnG1YG1YThrCYsVal1AsnGlnTrpAspAlaPheSerCYsGluCYs	1600
Db	3367	AACCTTGGCCACATATGGGGGACATTTGTGTGAACAGTGGACCGGTTCACTGTGAGATGC	3426

QY	1601	ProIeuNGIyPheGlyUGlyyUySseCYaAlaGInGluMetAlaAaPProGInhiSPhLeu	1620
Db	3427	CCCTCGGGCTTTGGGGGCGAAGAGCTGGGCCCAAGAAATGGCAATCCACAGACCTTCTCG	3486
QY	1621	GIySeSerSeIeuValAlaTPHhIGLyLeuSerIeuProIleSerGInProCTPTyLeu	1640
Db	3487	GGAGACAGCCTGGTGGCTGGCAAGGCTCTCGCTGCCATCTTCCAACTTGGTAACTCTC	3546
QY	1641	SeIeuMetPheATgThAArgGInAlaAspGlyValLeuLeuGInAlaIleThAArgGly	1660
Db	3547	AGCTCATGTTCCCAAGCGCCAGGCGCAAGGTGTCTGGCTGACAGGCATCCACCAAGGGGG	3606
QY	1661	ArgSeThriIeThLeuGInLeuAArgGInGlyhIstValMetIeuSeSerValGInGlyThr	1680
Db	3607	CGAGACACATCACTCTAACTAACTCAAGAGGGCCAGTAACTGAACCTGAGAGGGACACA	3666
QY	1681	GIyLeuGInAlaSeSerSeIeuAArgLeuGInProGlyAArgAlaAaAspGlyAAsPThhS	1700
Db	3667	GGGCTTCAAGGCTCTCTCTCTCGCTGAGCCAGGCGGGGCATATGACGGGTAACTGGCAC	3726
QY	1701	hIaAlaGInLeuAlaLeuGlyAlaSeSerGlyProGlyhIstAlaIleIeuSeSerPheAsp	1720
Db	3727	CATCAACAAGCTGGCACTGGAGGCCAGGGGGGGGCTGGCCCAATGCATTCGTCTTGCAT	3786
QY	1721	TyGlyGInGInAArgAlaGInGlyAAsPLeuGlyProAArgLeuhiSGIyLeuhiASeuSer	1740
Db	3787	TATGGCAGCAGAGAGCAGAGGGGCACTTGGGCCCGGGCTGCATGTTCTGGACTGAAC	3846
QY	1741	AaAIIeThValGlyGlyIleProGlyProAlaSGlyValAlaAArgGlyPheATgGly	1760
Db	3847	AACATAACAAGTGGCGGAAATACCTGGGCCAGCGCGGTGTGGCCCTGTGCTTCGGGGGC	3906
QY	1761	CyBLeuGInGlyValAArgValSeAAsPThrProGInGlyValAAsSeSerIeuAAsPProSer	1780
Db	3907	TGTTTGAAGGTGTGCGGGTGAAGGATACCGCAAGGGGGGTAAACAAGCTTGAATCCAGC	3966
QY	1781	hIaSGlyLeuSerIleAaAValGInGInGlyCySeSerIeuProAspProCYAAsPSeAAsPn	1800
Db	3967	CATGGGAGAGCAATCAACGTGGAGCAAGGCTGTAACTGTGCATCCTTGTGACTCAAC	4026
QY	1801	ProCyAsPProAlaAAsSeSerTyCySeSerAAsPThrAAsSeSerTySeSerCyAsP	1820
Db	4027	CCGATCTCTGTAAACAATACTGACAAAGCACTGGGAGCAAGCTATCTCTGAGCTGGAT	4086
QY	1821	ProGlyTyTyTyGlyAAsAAsPnCyvthAAsValCyAAsPLeuAAsPProCYvGIhIGIn	1840
Db	4087	CCAGGTAACTAATGGTGAACAATACTGAATATGTGTGTGAACCTGATGAGCAACAG	4146
QY	1841	SeValAlaCyvthAArgyAsPProSerAlaProhiSGlyTyThrCyvGIuCyvProProAAsP	1860
Db	4147	TCTGTGTGTACCCGCAAGCCCAAGTGGCCCCCAATGGCTAACTCTGGCAAGTGTCCCCCAAT	4206
QY	1861	TyTLeuGlyProTyCyvGIuThAArgIleAAsPGLProCyvProAArgIyTyThrPTyGly	1880
Db	4207	TACTTGGGCGCAATCTGTAGACCAAGATTGACCAAGCTTGTCCCTGGCTGGTGGGA	4266
QY	1881	hIaBProThrCyvGIyProCYvAAsPnCyAAsPValSeTySGlyPheAsPProAspCYvAAsPn	1900
Db	4267	CATCCCAACATGGGCCATCACTGAACCTGAGATGACAGAAAGCTTTGACCCAGACTGGAC	4326
QY	1901	LyvthSeSerGlyGInuCyvhiSGlyvGInuAAshiSTyThAArgProProGlySeSerProThr	1920
Db	4327	AAGACAAAGCGCGAGTGCACCTGCAAGAGAAACAACAACCGGCCCCCAAGGAGCGCCAC	4386
QY	1921	CyBLeuLeuCyAAsPnCyTyTyProThGlySeIeuSeSerAArgValCyAAsPProGInuAsP	1940
Db	4387	TGCTCTTGTGTAGACTGAACCCCAAGGCTCTTGTCCAGAGCTGTGACCCCTGAGAT	4446
QY	1941	GIyGInuCyvProCYvAsPProGlyValIleGlyAArgGInuCyAAsPArgCyAAsAAsPPro	1960
Db	4447	GGCCAGGTGCATGACCAAGCGAGGTATCTGGGCGGTGACGTACACGCTGTGACAACTCT	4506
QY	1961	PhehIaGInuValThThAAsGInuCYvGIuValAAsnTyAAsPSeCyvProAArgAlaIle	1980

Db	4507	TTTGCTGAGGTCAACCAACATGGCTGTAAGTAAGTAATATATGACAGTGGCCACGAGCGAATT	4566
Qy	1981	GLUALAGLYLLETPTPTProArgThzrPhegiyleuProalaaIAProCysePro	2000
Db	4567	GAGGTGGGAATCTGGTGGCCCCGTATCCGGCTTCGGGCTGCTGCTGCTCTCTGCTCC	4626
Qy	2001	LysegiyleSerPhegiylThrAlaValArgHisCyseAapgiuHisValrgiLYTPLeuProPro	2020
Db	4627	AAAGGCTCTCTTGGGACTGCTGTGGCCACTGTGATGACACAGGGGGTGGCTCCCCCA	4686
Qy	2021	AsnuLeuPheanCyLeThSerLleThrPheSergiuleuLysegiylPheAlaGLuArgLeu	2040
Db	4687	AACCTTCACACTGACAGCTGCATCACTTCCTCAGAACTGAAGGGCTTCGCTGAGGGCTCA	4746
Qy	2041	GLuArgaengiLuseSerLYLeuAapSergiYArgSerengiuLnuAlaLeuLeuLeuArg	2060
Db	4747	CAGCGGAATGAGTCAAGGCTGAGACTCAAGGGGCGCTCCACGACGCTAGCTGCTCTGGC	4806
Qy	2061	AenAlaThrgiLnuHisThAlaGLYTPhegiyleSerAapValLYseValAlaTYrgiLnu	2080
Db	4807	AACCCACGACGACACACAGCTGGCTACTTCGGACGACGACTCAAGAGTGGCTTACAGAGTG	4866
Qy	2081	AlaThzrArgLeuLeuAlaHisGLuseSerThrgiLnuArgiylPhegiyleuSerAlaThrgiL	2100
Db	4867	GCCACGGCGGCTGGCGCCACAGAGAGACCCAGCGGGGCTTGGGCTGTGCTGCACACAG	4926
Qy	2101	AapValHisPheThrgiLnuAsnuLeuArgValGLiseSerAlaLeuLeuAapThrAlaAen	2120
Db	4927	GACGTGCACCTTCACTGAGAAATCTGCTGGGGGTGGGACGGCTCTTGAACACAGCCACAC	4986
Qy	2121	LYsaArgHisThrgiLnuLeuileginiThrgiLnuArgiylThrAlaTYPLeuLeuGLuHis	2140
Db	4987	AAGCGGCACTGGAGAGCTGATCAGACAGACAGAGGGGTGCACCGCTGGCTGCTCCAGAC	5046
Qy	2141	TYrgiLnuAlaTYrAlaSerAlaLeuAlaGLuAapMetArgHisThrTYrLeuSerProPhe	2160
Db	5047	TATAGAGGCTTACGCAAGTGGCCCTGGCCACAGACGCGCACACCTTCAAGCCCCCTTC	5106
Qy	2161	ThriLeuAlaThrProAapniLeValileSerValAlaArgLeuAapLYseAapPheAla	2180
Db	5107	ACCATCTGCAGGCCCAACATTGTCTCTCCATGTGGCCCTTGGACAAAGGAACTTGTCT	5166
Qy	2181	GLyAlaLYseuProArgTYrgiLnuAlaLeuArgiylgiLnuinProProAapLeuGLuThr	2200
Db	5167	GGGGCCAAGTGGCCCGCTACAGAGGCCCTGGTGGGAGCAGGCCCGGACCTTGAGACA	5226
Qy	2201	ThrValilLeuProgiLnuSerValPheArgGLuThrProProValAlaArgProAlaGLY	2220
Db	5227	ACAGTCATTCGCGCTGAGTCTGTCTTCCAGAGAGCGCCCCCGTGTGTACGGCCCGCAGGC	5286
Qy	2221	ProgiLnuAlaGLuGLuProgiLnuLeuAlaArgArgGLuArgArgHisProgiLnu	2240
Db	5287	CCGGAGAGGCCCAAGAGCCAGAGAGGCTGGACGGCCACAGCCAGCGCACCCGGAGCTG	5346
Qy	2241	SergiLnuGLuAlaValAlaAsenValileileTYrArgThrLeuAlaGLYLeuLeuPro	2260
Db	5347	AGCCACGGGTGAGGGCTGTGGCCAGCGCTCATTCATCCGCACTTGGCGGGGCTACTGCT	5406
Qy	2261	HisenTYrAapProAapLYseArgSerLeuArgValProLYsaArgProileHisenThr	2280
Db	5407	CATATCTTTCAGCCCTGACAGCGCGCTGTGAGAGTCCCCAAACGCCGATCATCAACACA	5466
Qy	2281	ProValAlaSerLileSerValHisAapAapgiLnuLeuLeuProArgAlaLeuAapLYs	2300
Db	5467	CCCGTGTGAGCATCAGGCTCCATATATATATGAGAGCTTCTGGCCCCGGGGCCTTGACAA	5526
Qy	2301	ProValThrValGLuPheArgLeuLeuGLuThrgiLnuArgThLYsProileCyseVal	2320
Db	5527	CCCGTCAAGGTGACACTTCGGCTGTGGAGACAGAGCGGACCAACCCATCTGTCTC	5586
Qy	2321	PheThrAapniLuseLileuValSerGLYThrgiLnuLYTPSerAlaArgGLYCyseGLu	2340

Db	5587	TTCTGGAAACCATTCATCTCTGGTCAGTGGACAAGATGGAGCTGGCTGGCCACAGAGCTGTGAA	5646
Qy	2341	ValValPheaGAsnGluSerHisValSerCysGlnCysAsnHisSmetThrSerPheAla	2360
Db	5647	GTCTCTTCCGCAATGAGAGGCACGTACAGCTGCCAGTGGCAACACATACAGAGCTTGGCT	5706
Qy	2361	ValLeuMetAspValSerArgArgGluAsnGlyGluValLeuProLeuLysPheThrLeuThr	2380
Db	5707	GTGCTCATGGACGTTTCTCGCGCGGGAATGGGAGAGATCCCGACATGAAACACTGACA	5766
Qy	2381	TyrValAlaLeuGlyValAlaThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu	2400
Db	5767	TACGGGCTCTAAGAGTGCACCTTGGCTGGCCCTTGTGCTCACCTTCTCTCTCACTCTC	5826
Qy	2401	LeuArgIleLeuAspSerAsnGlnHisGlyValIleArgArgAsnLeuThrAlaAlaLeuGly	2420
Db	5827	TTGGGTATCCCTGGCCCTCCACACCAACAGAGGATCCGAGTAACTGACAGAGCTGCCCTGGGC	5886
Qy	2421	LeuAlaGlnLeuValPheLeuLeuGlyValLeuAsnGlnAlaAspLeuProPheAlaCysThr	2440
Db	5887	CTGGCTCAGCGTGGCTTCTCTCTGGGAATCAACAGAGCTGACCTCCCTTTGGCTGGACA	5946
Qy	2441	ValIleAlaIleLeuLeuHisPheLeuThrLeuGlyThrPheSerThrAlaLeuLeuGlu	2460
Db	5947	GTCAATGCATCCCTGGCTGCACCTTCTGTACTCTGCACCTTTTCTGGAGCTGTGCTGAG	6006
Qy	2461	AlaLeuHisIleLeuValArgAlaLeuThrGluValArgAspValAlaThrGlyProMetArg	2480
Db	6007	GCTTTCACCTGTATCCGGGACACTACAGAGGTGGCGAATGTCAACACCGGCCCATGCGC	6066
Qy	2481	PheTyrTyrrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Db	6067	TTCTACTACATCAGTGGGCTGGGCGTGGCTGCTCTTCAACAGGGCTAGCCGTGGGCTGG	6126
Qy	2501	AspProGlnGlyTyrGlyValAsnProAspPheCysPheLeuSerIleTyrAspThrLeuIle	2520
Db	6127	GACCCCGAGGAGCTACGGGAACCTTGACTTCTGTGGCTCTTCACATCAATGACACGCTATC	6186
Qy	2521	TyrrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu	2540
Db	6187	TGGAGTTTGTGTGGCCCGGTGGCTTTGGCCGTCTGCATGAGTGTCTTCTGTATCTCTG	6246
Qy	2541	AlaAlaArgAlaSerCysAlaAlaGlnArgGlnTyrrPheGluLysGlyProValSer	2560
Db	6247	GCGGCCCGGGCTCTGTGTCTGCCACAGCGGAGGGCTTGGAGAAAGGCTCTGTCTCG	6306
Qy	2561	GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrPheLeuAlaLeu	2580
Db	6307	GGCTTCAGGCTCTTCTTCCGCGTCTCTCTGTGTGTGAGCGCACGTGGCTGTGGCACTG	6366
Qy	2581	LeuSerValAlaSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle	2600
Db	6367	CTCTCTGTCAACAGGACACCCCTCTTCCACTCACTCTTTGTACTGTCAATTGCATC	6426
Qy	2601	GlnGlyProPheIlePheLeuSerTyrValValLeuSerTyrGluValArgValAlaLeu	2620
Db	6427	CAGGGCCCTTCATCTTCTCTCTTAATGTGGTGTGTAGCAAGAGAGTCCGGAAGCACTC	6486
Qy	2621	LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu	2640
Db	6487	AAGCTGTCTGCACGCCGCAAGCCCAAGCCCTGACCTCTGACACCAATTCACACCTGC	6546
Qy	2641	ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly	2660
Db	6547	ACCTGTCTTCAACTGCCCCCAAGCCCTTAGCGAATGGCGGCTGTACAGCCTTAGCGA	6606
Qy	2661	AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle	2680
Db	6607	GACTCGGCCGAGCTCTGTGCACAGCACAGTGTGTGGGCAAGAGTCAAGCCCAAGCTACATC	6666
Qy	2681	ProPheLeuLeuArgGlnGluSerValAlaLeuAspProGlyGlnGlyProProGlyLeuGly	2700
Db	6667	CCCTTCTTGTGTGAGAGAGAGTGTGGCACTGAACTTGGCCAAAGGAGCCCTCTGGCTGGGG	6726





Db	533	GGGACCTATGCAATGATGAGTGGCAATGCTCGGGACAGTTTATCTGGATCCCAACT	602
OY	454	GjYalaleuAepValValSerProleuAepTyrGluThrThryleGluTyrThryleuArg	473
Db	603	GGAGCTCTGATGTGTGGTGGCCCTCTTGACTATGAGACCAAGAGTACCCCTTACGG	662
OY	474	ValAaGlaagInAepGlyGlyAArgProleuSerAepValSerGlyLeuValThrVal	493
Db	663	GTGGCAGACAGAGATGTGGCCGTCCGCCACTGTAAATGTCTGGCTTGGTGTGACAGTA	722
OY	494	GlnValleuAepTlAepAepAepAaProlepeValSerThrProPheGlnAaThr	513
Db	723	CAGGTCTCGAATATCAAGCAATAGTCCCACTTGTGTGAGACCCCTTCCAGGCTACT	782
OY	514	ValleuGluSerValProleuGlyTyrLeuValleuHisValGlnAaIleAepAaAep	533
Db	783	GTCTCGAAGAGTCCCTTACGCTACCTGGTCTTCATGTCAGAGCTATGACACCTGAT	842
OY	534	AlaGlyAepAepAaAArgLeuGluTyrAArgleuAaGlyValGlyHisAepPheProPhe	553
Db	843	GCTGTGTGCAAAAGCCCGCTGGAAATACCGCTTGTGGGTGGGACATGACTTCCCTTC	902
OY	554	ThrIleAepAepGlyThrGlyTyrPrlleSerValAaIaGluLeuAepAArgGlnVal	573
Db	903	ACCATCAACAATGGGACAGAGCTGGAGTCTTGTGGCTGAACTGACCGGGAGGAAGTT	962
OY	574	AepPheTyrSerPheGlyValGluAaArgAepHisGlyThrProAlAlepThrAlaSer	593
Db	963	GATTTCTACAGCTTTGGGTAGAACTCGAAGCAATGACATCTCCAGCACTCACTGCTCG	1022
OY	594	AlaSerValSerValThrValleuAepValAepAepAepAepProThrPheThrGlnPro	613
Db	1023	GCCAGTGTCAAGGTGACTGTCTCGATGTGCAACCAATCCAACTTTATCCACCA	1082
OY	614	GluTyrThrValAArgLeuAepGluAepAaIaValaGlyThrSerValValThrValSer	633
Db	1083	GAGTACACAGTGGCGGCTCAATAGATGACAGCTGTGGGACCAAGGTGTGACGGTGTCA	1142
OY	634	AlaValAepAArgAepAaHisSerValIleThrTyrGlnIleThrSerGlyAepAThrArg	653
Db	1143	GCTGTGGACCGGTAGCTCATAGTGTCACTCACTTACAGATACCAATGGGCAATCTGCA	1202
OY	654	AepAArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaAepProleu	673
Db	1203	AACCGCTTCTCATCACCAGCCAAAGTGGTGGCTGTATCTCCCTTGCCCTGACATG	1262
OY	674	AepTyrTylleuGluAArgGlnTyrValleuAlaValThrAlaSerAepGlyThrArgGln	693
Db	1263	GACTTCAAACTTGACCGGACAGTATGTGTTGGCTGTACCGCTCCGATGGCACTCGGCA	1322
OY	694	AepThrAlaGlnIleValValAepValThrAepAaAepThrHisAArgProValPheGln	713
Db	1323	GACATCGGACAGATGTGGTGTGATGTACCCAGCGCAACCACTGATGCTGTCTTTGAG	1382
OY	714	SerSerHisTyrTyrThrValAepValAepGluAepAArgProAlaGlyThrThrValValleu	733
Db	1383	AGCTCCCACTAATACAGTGAATTTATAGAGACCGGCGGCAAGGACCAACGGTGGTGTG	1442
OY	734	IleSerAlaThrAepGluAepThrGlyGluAepAaArgIleThrTyrPheMetGluAep	753
Db	1443	ATCAGCGCCACGATGAGGACCAAGTGTGAATGCCCATCATCACTTTCATGAGAGAC	1502
OY	754	SerIleProGlnPheArgIleAepAaAepThrArgIlaValThrThrGlnAlaGluLeu	773
Db	1503	AGCATATCCCAAGTTCGATCGATGATGACAGCACGGGGGCTGTACCAACCGAGCTAGCTG	1562
OY	774	AepTyrGluAepGlnValSerTyrThrIleuAlaIleThrAlaArgAepAepGlyIlePro	793
Db	1563	GACTACGAAGACCAAGTGTCTTACACCTGTGGCAATTACTGCTCGGGACAAATGGCAATTCC	1622
OY	794	GlnIySerSerAepThrThrTyrLeuGluIleleuValAepAepValAepAepAepAaPro	813

Db	1623	CAGAAAGTCCGACACCACTTACCTTGAGATCTCTGGTGAACGACGTGAATGACATAGCCCTT	1682
Qy	814	GIhPheLeuAaGspSerTYrGInGIySerValTYrGIuaspValProProPheThSer	833
Db	1683	CAGTTCCGTGGAGACTCTTACCAAGGGAGTGTCTATAGAGATGTGGCCACTTCACTAGC	1742
Qy	834	ValLeuGInIIeSerAlaThrAspArgAspSerGIyLeuAsnGIyAArgValPheTYrThr	853
Db	1743	GTCCCTGAGACTCTAGGCCACTGATCTGATTTCTGACCTTAATGACAGGCTCTTCAACCC	1802
Qy	854	PheGInGIyGIyAAspAspGIyAAspGIyAAspPheIleValGIuSerThrSerGIyIleVal	873
Db	1803	TTCCAGAGAGGCGAGAGAGAGAGAGAGAGTCTTATTTGTAGATCCAGTCACGGATCGTG	1862
Qy	874	ArgThrLeuAaGArgLeuAspArgGIuAsnValAlaGInTYrValIleAaArgAlaTYrAla	893
Db	1863	CGAAGCGTACGAGAGGCTGAGATCGAGAGACCTGTGCCACGATGTGCTTGCGGGCATATGCA	1922
Qy	894	ValAspIySGIyMetProProAlaAArgThrProMetGIuValIThrValIleuAsp	913
Db	1923	GTGGCAAGGGGATGCCGCCGACCCGACACACTATGAGAAAGTGAACATCATCTGTGGAT	1982
Qy	914	ValAspAspAsnProProValAlaPheGIuGIuAspGIuPheAspValPheValGIuGIuAsn	933
Db	1983	GTGATATGCAATCCCCCTGTCTTTGAGCAGAGATGAGTTGATGTGTGTGGAGAGAAC	2042
Qy	934	SerProIleGIyLeuAlaValAlaAArgValIThrAlaThrAspProAspGIuGIyThrAsn	953
Db	2043	AGCCCATTTGGGCTTAGCCGTGTGGCCGGGTCAAGCCACTAACCCTGATGAAGGCAACAT	2102
Qy	954	AlaGInIIeMetTYrGInIleValGIuGIyAsnIleProGIuValPheGInLeuAspIle	973
Db	2103	GCCAGATTTATGTACAGATTTGTGAGGGGCAACATCCCTGAGGTCTTCAGCTGGACATC	2162
Qy	974	PheSerGIyGIuLeuThrAlaLeuValAspLeuAspTYrGIuAspArgProGIuTYrVal	993
Db	2163	TTCTCCGGGGAGCTGACAGCCCTGTGTGACTTGTAGACTACAGAGACCGGCTGTAGTACGT	2222
Qy	994	LeuValIIleGInAlaThrSerAlaProLeuValSerArgAlaThrValIHisValArgLeu	1013
Db	2223	CTGTGATTCACGAGCCACAGTCAAGTCTCTCTGTGTGAGCCGGGCTTACAGTCCAGTCCGCTC	2282
Qy	1014	LeuAspArgAsnAspAsnProProValIleuGIyAsnPheGIuIleLeu	1029
Db	2283	CTTGACCGCATATGACCAACCAACGATCTGTGGGCAACTTTAGATTCCTT	2330
RESULT 9			
AAK83060			
ID	AAK83060	standard; DNA; 2332 BP.	
XX	AC		
XX	AAK83060;		
DT	07-NOV-2001	(first entry)	
XX			
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37872.		
XX			
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157182-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01354.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184654.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		







QY 854 PheGlnGlyGlyAspAspGlyAspPheIleValIuSerThrSergIylIleVal 873  
DB 1803 TTCGAGGAGGCGGACGATGAGACGGTGACTTATTTGTCACGTCAGGACATCGTG 1862  
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnIlyValLeuAlaIyTAla 893  
DB 1863 CGAACGCTACGGAGGCTGGATCGAGAACGTGGCCGAGTATCTTGGCGGATATGCA 1922  
QY 894 ValAspIleGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913  
DB 1923 GTGGACAGGAGGATGCCCCAGCCCGCACCTATGGAACTGACAGCTGCTTGCAAT 1982  
QY 914 ValAsnAspAsnProProValPheGluGlnAspGlnPheAspValPheValGluGluAsn 933  
DB 1983 GTCAATGACATATCCCCCTCTTGTGACAGATGAGTTGATGTGTTGTGGAAGAGAAC 2042  
QY 934 SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn 953  
DB 2043 AGCCCATTTGGGCTAGCGGTGGCCGGGTCAAGCCCATGACCCCGATGAGGACCAAT 2102  
QY 954 AlaGlnIleMetIlyGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIle 973  
DB 2103 GCCCATATTATATGACGATTTGTGAGGGGACATCCCTAGGCTTCCAGCTGGACATC 2162  
QY 974 PheSergIyGlyLeuThrAlaLeuValAspLeuAspIlyGluAspArgProGluIlyVal 993  
DB 2163 TTCTCCGGGAGCTGACAGCCCTGTGATGACTTACGACGAGGACGGCCTGAGTACGTC 2222  
QY 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValaGluLeu 1013  
DB 2223 CTGGTATCCAGGCGCAGTACGCTCCTGTGTAGACGGGCTACAGTCCAGTCCGCTTC 2282  
QY 1014 LeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeu 1029  
DB 2283 CTGGACCGGATGACCAACCAACGTCGTGGCACTTTGAGATCCTT 2330

RESULT 10  
AAS31490  
ID AAS31490 standard; DNA; 2332 BP.  
XX  
AC AAS31490;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human DNA for a novel extracellular matrix protein, Seq ID No 569.  
XX  
KM Human; secreted extracellular matrix protein; ds; immunomodulatory;  
XX Anti-HIV; antineoplastic; antirheumatic; antisclerotic; cardiant; vascular;  
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
XX antialzheimers; immune/autoimmune disease; HIV infection; anaemia;  
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
XX cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
XX Sezary syndrome; Gaucher's disease; neurological diseases;  
KM Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KM cardiac arrest; tachycardia; angina; infection; corneal infections;  
KM wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO200155368-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01348.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0215647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234999.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465572/50.  
XX  
XX Nucleic acid molecules encoding human secreted extracellular matrix  
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
XX  
XX Claim 1; SEQ ID No 569; 577bp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules encoding  
CC novel human secreted extracellular matrix proteins (SPs). The  
CC polynucleotides and proteins are used to prevent, treat a medical  
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,

CC chickens or sheep. For example, disorders associated with decreased  
CC expression of SPs. The SP polynucleotide or a vector expressing them may  
CC be administered to treat diseases by gene therapy. Antisense molecules  
CC may be administered to down regulate expression of SPs by binding with  
CC the cells own genes and preventing their expression. The polynucleotides  
CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
CC used as antigens to produce antibodies and to identify modulators  
CC (agonists and antagonists) of the SPs. The anti- (SP) antibodies and  
CC antagonists may also be used to down regulate expression and activity of  
CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anemia, rheumatoid arthritis  
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
CC include wound healing, maintenance of organs before transplantation,  
CC support of cell culture of primary tissues, modulation of for example

Alignment Scores:

Pred. No.: 0 Length: 2332  
Score: 776.00 Matches: 776  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.55% Indels: 0  
DB: 22 Gaps: 0

US-09-916-849a-3 (1-2923) x AAS31490 (1-2332)

QY 254 ThrHisValPheargValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
DB 3 ACCACGCTTTCAGGGTCACGGCCGACAGACCAAGGATGCGCTGGCT 62  
QY 274 ThrLeuThrIleLeuValThrAspThrAspHisAspProValPheGlnGlnGlu 293  
DB 63 ACACCTACCAATCTTGTTACTGACACCAATACATACCTCTGTTGAGACAGAG 122  
QY 294 TyrIleGlnSerLeuAspGluAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
DB 123 TACAAGAGAGCCCTCAGGAGAGAACCTGAGGTTGCTATAGAGTCTACGTCAGGGCC 182  
QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlySerGly 333  
DB 183 ACGGATGGTGAATGCCCTCCCAATGCAATATCTGTAACCGCTGGAGGGCTGGG 242  
QY 334 GlySerProSerGluValPheGlnIleAspProArgSerGlyValIleArgThrArgGly 353  
DB 243 GGCAGCCCTCTGAAGCTTTGAGATGACCTGCTGCGGGTGAATCCGAGC 302  
QY 354 ProValAspArgGlnGluValGlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGly 373  
DB 303 CCTTGATCGGAGAGAGGTGATCTTCAAGTGAAGGAGGCAAGTGAACCAAGGT 362  
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393  
DB 363 CGGAGCCCGGGTCTCGAAGTACCAAGCCGCTGTTTCTTCTCGTGAGGATGACAT 422  
QY 394 AspAsnAlaProGlnPheSerGlnIleValArgTyrValValGlnValArgGluAspValThr 413  
DB 423 GATATAGCCCCCGATTAGTGAAGAGCGCTATGTGCTCAGTGAAGAGAGATGACT 482  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
DB 483 CCAAGGGGCCAGTACTCCAGTCAAGCTCGGATGAGACAGAGGAGCAAGCCGCTG 542  
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
DB 543 GTGCACTATAGCAATCAGAGTGGCAATGCTCGGGAGCAAGTTTATCTGAGGCCAGACT 602  
QY 454 GlyAlaLeuAspValValIleSerProLeuAspTyrGlnThrThrIleGlyGluTyrThrLeuArg 473

Db 603 GGAAGCTCTGGATGTGTGAGCCCTCTTGAATATAGACGACCAAGAGTACACCCCTACG 662  
 Qy 474 ValArgAlaGlnAspGlyIa-GrProLeuSerAsnValSerGlyLeuValThrVal 493  
 Db 663 GTCCGAGCAAGAGATGGTGGCCGCTCCCACTCTTAATCTCTGGCTTGGAGCAAGTA 722  
 Qy 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
 Db 723 CAGGCTCTGGATATCAACGACCAATGGCCCCCATCTTGGTCAGCAACCCCTTCCAGGCTACT 782  
 Qy 514 ValLeuGlnSerValProLeuGlyValLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
 Db 783 GTCTGGAGAGGCGTCCCTTAAGCTACCTGGTTCTCCATATGCCAGGCTATCGACGCTGAT 842  
 Qy 534 AlaGlyAspAsnAlaArgLeuGlyValArgLeuAlaGlyValGlyHisAspPheProPhe 553  
 Db 843 GCTGGAGCAATGCCCGCTGGAAATACCGCTTGGTGGGTGGAGCATGACTTCCCTTC 902  
 Qy 554 ThrIleAsnAsnGlyThrGlyValProIleSerValAlaAlaGlyLeuAspArgGlyVal 573  
 Db 903 ACATCAACAAATGGACAGAGCTGATCTCTGTGGCTGACCTGACCTGGACCGGAGGAAGTT 962  
 Qy 574 AspPheTyrSerPheGlyValGlyAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
 Db 963 GATTTCTACAGCTTTGGGGTAGAAGCTCGAGACCAATGGCACTCGACGACTCATGCTCTG 1022  
 Qy 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613  
 Db 1023 GCCAGGTGACGCGTGAAGCTGCTGGATGTCAACGACCAATCCAACTTTACCCACCA 1082  
 Qy 614 GlnTyrThrValArgLeuAsnGlyAspAlaAlaValGlyThrSerValValThrValSer 633  
 Db 1083 GACTTACACAGTGGGCTCATGTAGATGAGTGCAGCTGTGGCACCGCGTGGAGCGGTGTCA 1142  
 Qy 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653  
 Db 1143 GCTGTGACCGCTAATGCTCATAGTGTATCACTTACCAATTCACAGTGGCAATACTCA 1202  
 Qy 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673  
 Db 1203 AACCGCTTCCATTCACACGACCAAGGTGGTGGCTGATCCCTTGGCCCTGACACG 1262  
 Qy 674 AspTyrIleLeuGlnIleThrGlnIleValIleValIleValThrAlaSerAspGlyThrArgGln 693  
 Db 1263 GACTTACAAACTTGAAGCGGAGTATGTGTGCTGTATCCGCTCCCAATGCGACTCGGCA 1322  
 Qy 694 AspThrAlaGlnIleValIleAsnValThrAspAlaAsnThrHisArgProValPheGln 713  
 Db 1323 GACACGGCACAAGTGTGGTGAATGTCAACGACGCAACCAATGCTCTGCTTTCCAG 1382  
 Qy 714 SerSerHisTyrThrValAsnValAsnGlyAspArgProAlaGlyThrThrValValLeu 733  
 Db 1383 AGCTCCCACTATATACAGTAATGTTAATAGAGGACCGGCGGACGACCAAGGTGGTCTG 1442  
 Qy 734 IleSerIleAlaThrAspGlyAspThrGlyLeuAsnAlaArgIleThrTyrPheMetGlyAsp 753  
 Db 1443 ATGAGGCGCCAGATGAGGACACAGGTGAAATAGCCGCAATCACTTCAATGAGGAGAC 1502  
 Qy 754 SerIleProGlnPheArgIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGlyLeu 773  
 Db 1503 AGCATCCCGCCAGTTCGCCATCGATGACGACACCGGGGGCTCTCAACCCAGCTGAGCTG 1562  
 Qy 774 AspTyrGlyAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
 Db 1563 GACTACCAACACCAAGTGTCTTACACCTGCGCATTAAGCTCTGGGCAATGACATTTCCC 1622  
 Qy 794 GlnIleSerAspThrThrTyrIleGlnIleLeuValAsnAspValAsnAspAsnAlaPro 813  
 Db 1623 CAGAGTCCCAACACCACTACCTGAGATCTGGTGAACAGCTTAATGACATGCCCCCT 1682  
 Qy 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGlyLeuAspValProPheThrSer 833  
 Db 1683 CAGTCTCTGGAGACTCTCTACAGGGGACGTGTCTATAGAGATGTGGCAACCTTCACTAGC 1742

Qy 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853  
 Db 1743 GCTCTGAGATTCACGCCACTGATTCGTGAATTAATGGAGGGTCTTCAACAC 1802  
 Qy 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGlySerThrSerGlyIleVal 873  
 Db 1803 TTCCAAAGAGGGGAGAGATGAGACCGGTATTTATGTGTAGTCCACGTCAGGCAATCGTG 1862  
 Qy 874 ArgThrLeuArgArgLeuAspArgGlyAsnValAlaGlnIleValIleValLeuArgAlaTyrAla 893  
 Db 1863 CGAAGCTACGAGGCTGATGTGAGAGAACGTGGCCAGTATGTCTTGGGGCATATGCA 1922  
 Qy 894 ValAspIleGlyMetProProAlaArgThrProMetGlyValThrValThrValLeuAsp 913  
 Db 1923 GTGACCAAGGGGATGCCCGCCAGCCGACCACTTAAGAAATGACAGTCACTGTGTTGAT 1982  
 Qy 914 ValAsnAspAsnProProValPheGlyGlnAspGlyPheAspValPheValGlyLeuAsn 933  
 Db 1983 GTGAATGACAAATCCCGCTGTCTTGAAGAGATGATGTGTGTGTGAGAGAGAAC 2042  
 Qy 934 SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlyIleThrAsn 953  
 Db 2043 ACCCCCATTTGGCTTACCGCTGGCCCGGTCAAGCCCACTGACCCCGATGAGGACCAAT 2102  
 Qy 954 AlaGlnIleMetTyrGlnIleValGlyLeuGlyAsnIleProGlyValPheGlnLeuAspIle 973  
 Db 2103 GCCAGATTAATATACAGATTTGTGAGGGGCAACATCCCTGAGGTCTTCCAGCTGACATC 2162  
 Qy 974 PheSerGlyGlyLeuThrThrAlaLeuValAspLeuAspTyrGlyAspArgProGlyIleVal 993  
 Db 2163 TTCTCCGGGAGACTTACAGCCCTGTGTAGACTTACAGAGAGACCGGCTGAGTACGTC 2222  
 Qy 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1013  
 Db 2223 CTGTGTATCCAGGCCACAGCTGACTCTCTGTGTGAGCCGGGCTACAGTCCAGCTCCGCTC 2282  
 Qy 1014 LeuAspArgAsnAspAsnProProValLeuGlyAsnPheGlnIleLeu 1029  
 Db 2283 CTTGACCCCAATGACCAACCCACCACTGCGGGCACTTGTAGATCCTT 2330  
 RESULT 11  
 ABO66814  
 ID ABO66814 standard; DNA; 2332 BP.  
 XX  
 AC ABO66814;  
 XX  
 DT 23-AUG-2002 (first entry)  
 XX  
 DX Human polynucleotide SEQ ID NO 569.  
 XX  
 KW Human; noctropic; neuroprotective; cytoskeletal; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antischizoid; antianemic; antidiabetic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antitumor;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX  
 XX US2002042386-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 17-JAN-2001; 2001US-0764870.  
 XX  
 PR 31-JAN-2000; 2000US-179065P.  
 PR 04-FEB-2000; 2000US-180628P.  
 PR 28-JUN-2000; 2000US-214886P.  
 PR 07-JUL-2000; 2000US-216647P.  
 PR 07-JUL-2000; 2000US-216880P.  
 PR 11-JUL-2000; 2000US-217487P.

PR 11-JUL-2000; 2000US-217496P.  
 PR 14-JUL-2000; 2000US-218290P.  
 PR 26-JUL-2000; 2000US-220633P.  
 PR 26-JUL-2000; 2000US-220964P.  
 PR 14-AUG-2000; 2000US-224518P.  
 PR 14-AUG-2000; 2000US-224519P.  
 PR 14-AUG-2000; 2000US-225267P.  
 PR 14-AUG-2000; 2000US-225268P.  
 PR 14-AUG-2000; 2000US-225270P.  
 PR 14-AUG-2000; 2000US-225477P.  
 PR 14-AUG-2000; 2000US-225757P.  
 PR 14-AUG-2000; 2000US-225758P.  
 PR 22-AUG-2000; 2000US-226868P.  
 PR 30-AUG-2000; 2000US-228924P.  
 PR 01-SEP-2000; 2000US-229287P.  
 PR 01-SEP-2000; 2000US-229343P.  
 PR 01-SEP-2000; 2000US-229344P.  
 PR 01-SEP-2000; 2000US-229345P.  
 PR 05-SEP-2000; 2000US-229509P.  
 PR 05-SEP-2000; 2000US-229513P.  
 PR 08-SEP-2000; 2000US-231413P.  
 PR 21-SEP-2000; 2000US-234233P.  
 PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236327P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-237045P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 PR 08-DEC-2000; 2000US-251869P.  
 PA (ROSE/) ROSEN C A.  
 PA (RUBB/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 DR WPI: 2002-470713/50.  
 XX  
 PT New nucleic acid encoding human proteins, useful for diagnosis,  
 PT treatment and prevention of e.g. osteoporosis, also related  
 PT polypeptides and antibodies -  
 XX  
 ES Disclosure; SEQ ID NO 569; 235bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (AB06521-AB06785) and proteins  
 CC (AB04746-AB04810) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections. The present sequence is that of a  
 CC polynucleotide of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from USPTO at [seqdata.uspto.gov/sequence.html?docID=99909764870](http://seqdata.uspto.gov/sequence.html?docID=99909764870).  
 XX  
 SQ Sequence 2332 BP, 505 A; 669 C; 665 G; 493 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2332  
 Score: 776.00 Matches: 776  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.55% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-916-849a-3 (1-2923) X AB066814 (1-2332)  
 QY 254 ThrHsValPheArgValThrAlaGlnAspHisGlyMetProAlaGlySerAlaLeuAla 273  
 DB 3 ACCACGCTTCAGAGTCACGCGCAGACCAACGCGCATGCCCAAGAGTGCCTGCT 62  
 QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlu 293  
 DB 63 ACACCTCACCATCTTGTTACTGACACCATATACATACCTGTGTTCAGACAGCAGG 122  
 QY 294 TyrIysGluSerIleuArgIuAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAGAGAGAGCTCAGAGGAACTCAGAGTTGGCTATAGAGTGCTCATGTCAGGGCC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuGluGlySerGly 333  
 DB 183 ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 242  
 QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GCGAGCCCTCTGAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 302  
 QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
 DB 303 CCTGTGATCGGAGAGAGTGAATCCTACACGATGAGGAGGAGGAGGAGGAGGAGG 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaValPheLeuSerValGluAspAsn 393  
 DB 363 CGGAGCCCGGCTCTGAGAGTGAATCCTACACGATGAGGAGGAGGAGGAGGAGGAG 422  
 QY 394 AspAsnAlaProGlnPheSerGluValArgTyrValGluValArgGluAspValThr 413  
 DB 423 GATATATGCCCCCGATTTAGTGAAGACCGCTATGTGTCAGAGGAGGAGATGTACT 482  
 QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
 DB 483 CCAGGGGCCCCAGATCTCCGATCAGACCTCGATGAGGAGGAGGAGGAGGAGGAGG 542  
 QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
 DB 543 GTGCACATAGACATCAAGATGAGTGCAGATGCTCGGGAGAGATTATATCGAGTCCCAACT 602  
 QY 454 GlnValLeuAspValValSerProLeuAspTyrGlnThrThrIysGluTyrThrLeuArg 473  
 DB 603 GGACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 662  
 QY 474 ValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
 DB 663 GTGGAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 722  
 QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
 DB 723 CAGGTCTGATATCAACGACATGCCCCCATCTTCGACGACCCCTTCAGGGTACT 782  
 QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
 DB 783 GTCTTGAGAGAGCGTCCCTTAGGTACTGTTCTTCATATGACGAGGTTATGACGTGAT 842



OY	534	AlaGlyAspAspAlaAArgLeuGluYrYrArgLeuAlaGlyValAlaGlyHisAspPheProHe	553
Db	843	GCTGGTGCAMATGCCCGCTGGAAATACCGCTTGTCTGGGGGTGGACATGATCTTCCCTTC	902
OY	554	ThrIleAsnAnglyYrThrGlyYrTrpIleSerValAlaAlaGlyLeuAspArgGluGluVal	573
Db	903	ACCATCAACAAATGGCAGAGCTGGATCTCTGGCTGTGAACCTGACCCGGAGGAAGTT	962
OY	574	AspPheYrSerPheGluValGluAlaArgAspHisGlyYrThrProAlaLeuThrAlaSer	593
Db	963	GATTTCTACAGCTTTGGGTGAGAGCTGAGACCAATGGCACTCCAGCACTACGCTCCG	1022
OY	594	AlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGluPro	613
Db	1023	GCCAGGTGCAGGTGATCTGCTCTGGATGTCAACGACAAACAACTCCACTTTACCAACCA	1082
OY	614	GluYrYrThrValAlaGlyLeuAsnGluAspAlaAlaValGlyYrThrSerValThrValSer	633
Db	1083	GAGTACACAGTGGCGGCTCATATGAGGATCAGCTGTGGGACACAGGTGGTACAGGTGCA	1142
OY	634	AlaValAspArgAspAlaHisSerValIleThrYrGlnIleThrSerGlyAsnThrArg	653
Db	1143	GCTGTGGACCGGTGATGCTATAGTGTATCACTACCTACAGATCAACAGTGGCAATCTGCA	1202
OY	654	AsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	1203	AACCGCTTCTGCATCAACAGCCAAAGTGGTGGGTGGTATCCCTTGCCCTGCACGTG	1262
OY	674	AspYrYrIleLeuGluArgGlnIleYrThrValLeuAlaValThrAlaSerAspGlyYrThrArgGln	693
Db	1263	GACTACAAACTTGAGCGGAGATGTGTGGTGTGTACCGCTCCGATGGCACTCGGAG	1322
OY	694	AspThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	1323	GACACGGCACAGATTGTGTGTGATGTCAACCGCACACCCACCGTCTGTCTTTCAG	1382
OY	714	SerSerHisYrYrThrValAsnValAsnGluAspArgProAlaGlyYrThrValValLeu	733
Db	1383	AGCTCCCACTATACAGTAAATGTATATAGAGCCGGCGGCGGACACACGCTGTGTCTG	1442
OY	734	IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrYrPheMetGluAsp	753
Db	1443	ATCAGCGGCACGAGTGAAGACACAGGTGAGATGCCCGCATCACTTCTCATGGAGAC	1502
OY	754	SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu	773
Db	1503	AGCATCCCCCAGTGTCCGATCGATGCAGACACGGGGGCTGTCAACCAAGCTGAGCTG	1562
OY	774	AspYrYrGluAspGlnValSerYrYrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	1563	GACTACGAAGACCAAGTGTCTTACACCCCTGGCATTTACTGCTCGGAGACAAATGGCAATTC	1622
OY	794	GlnYrSerAspThrYrYrLeuAluIleLeuValAsnAspValAsnAspAsnAlaPro	813
Db	1623	CAGAGTCCGACACACCACTTACCTGAGATCCGTGGTGAACAGACTAAATGACATGCCCT	1682
OY	814	GlnPheLeuArgAspSerYrGlnGlySerValYrGluAspValProProPheThrSer	833
Db	1683	CAGTTCCTGGCGAGCTCTCCACCAAGGCGAGTGTCTTATGAGAGTGTCCACCCCTTCACTAGC	1742
OY	834	ValIleAsnGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyAlaArgValPheYrYrThr	853
Db	1743	GTCCTCGAGATTCAGCCACTAATGTGATTTGAGACTTAATGGACGGGGTCTTCTTACACC	1802
OY	854	PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal	873
Db	1803	TTCCAAAGAGGCGAGATGAGACGGGTGACTTTATTTGTTGAATGTCCACGTCAGGCATCGTG	1862
OY	874	ArgThrIleuArgArgLeuAspArgGluAsnValAlaGlnYrValLeuArgAlaYrYrAla	893
Db	1863	CGAAGCGTACGAGAGGTGATGAGAGAACTGTGGCCCAAGTATGTCCTTGGGGCATATGTGA	1922

Oy		894	ValAspApysGlyMetProPcoAlaaArgThrPromedGlValThrValThrValIleuAsp	913
Dd		1923	GtGAACAAGGGGATGCCCCCAAGCCCCGACACTHrTGAATGACAGTCGTGGAT	1982
Oy		914	ValAsnAspAnpProproValPheglugInaspGluPheAspValPheValGlugluAsn	933
Dd		1983	GtGAATGACATCCCTCGTCTTTGAGCAGAGTGATTGATGTGTGGAAAGAAC	2042
Oy		934	SerProIleglueuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn	953
Dd		2043	AGCCCCATGGGCTAGCGCGTGGCCCCGGGTCAAGCCAATAACCCGATGAAGGACCAAT	2102
Oy		954	AlaGlnIleMetTyrglnIleValGluglyAsnIleProGluValPheGlnLeuAspIle	973
Dd		2103	GCCCAAGTATTATGCACGATTGTGAGAGGACCATCCCTGAGGTCCTTCCAGCTGCACATC	2162
Oy		974	PheSerelYglIleuThrAlaLeuValAspLeuAspTyrgluAspAqProGluTyrrVal	993
Dd		2163	TTCCTCGGGAGCTGACAGCCCTGGTAGACTTGACTTGAACACAGACCGGCTGAGTAAGTC	2222
Oy		994	LeuValIleglnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu	1013
Dd		2223	CTGTCTATCCAGGCGCACGTCAAGCTCCCTGTGTAGCCGGGCTACAGTCCAGTCCGCTC	2282
Oy		1014	LeuAspArghAsnAspAnpProproValIleuGlyAsnPhelGluIleu	1029
Dd		2283	CTTGACCGCAATGACAAACCACCAAGTGTGGGGCAACTTGAAGATCCCTT	2330
<b>RESULT 12</b>				
ID	ABTJ1943		standard; DNA; 2391 BP.	
XX	ABTJ1943;			
AC				
DT	01-MAY-2003	(first entry)		
DE			Human breast cancer / ovarian cancer related coding sequence #50.	
XX				
KW			Human; gene; ds; cytoabatic; breast cancer; ovarian cancer.	
XX				
OS			Homo sapiens.	
XX				
PN	WO2003000012-A2.			
PD	03-JAN-2003.			
PF				
PP	21-JUN-2002; 2002WO-US19773.			
PR	21-JUN-2001; 2001US-300159P.			
PR	27-JUN-2001; 2001US-301351P.			
PA	(MILL-) MILLENNIUM PHARM INC.			
PI	Velby OP;			
DR	MP1; 2003-267848/26.			
DR	P-BSDb; ABJ37074.			
PT			Determining the presence of breast cancer in an individual, involves	
XX			using specific polynucleotide markers -	
PS			Disclosure; Page 218; 233pp; English.	
XX			The invention comprises a method for assessing whether a patient is	
CC			afflicted with breast cancer or ovarian cancer. The method involves the	
CC			use of specific DNA markers. The method of the invention is useful in the	
CC			detection and treatment of ovarian and breast cancer. DNA sequences	
CC			ABTJ1894 - ABTJ1949 encode human breast/ovarian cancer-related proteins.	
XX				
SO			Sequence 2391 BP; 460 A; 810 C; 679 G; 442 T; 0 other;	
<b>Alignment Scores:</b>				
Pred. No.:	0		Length:	2391

Score: 729.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.94% Indels: 0  
DB: 25 Gaps: 0  
US-09-916-849a-3 (1-2923) x ABT31943 (1-2391)  
QY 2186 ArgTGTGUAAlaLeuArgGlyGluGlnProProAspLeuGluThrThrValIleLeuPro 2205  
DB 19 CGCTACGAGGCCCTGGGTGGAGGAGAGCCCGGAGCTTGAAGACAAAGCTATTCGCTT 78  
QY 2206 GluSerValPheArgGluThrProProValValArgProAlaGlyProGlyGluAlaGln 2225  
DB 79 GAGTCTGTCTTCGAGAGAGAGCCCGGTGTCTAGAGCCCGGAGGCCCGGAGGCCAG 138  
QY 2226 GluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGlyGluAla 2245  
DB 139 GAGCCAGAGAGAGCTGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198  
QY 2246 ValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspPro 2265  
DB 199 GTGGCCAGGCTATCATCTACCGCACCTGGCCGGCTACTGCTCATTAATAGACCTT 258  
QY 2266 AspLysArgSerLeuArgValProLysArgProIleIleAsnThrProValValSerIle 2285  
DB 259 GACAGCCGAGCTTGGAGAGTCCCAACGCCGATCATCAACACCCGTGGTGAAGCATC 318  
QY 2286 SerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysProValThrValGln 2305  
DB 319 AGGCTCATAGATGATAGAGAGGCTTCTGCCCGGAGCCCTGAGCAAAACCCGTCAAGTGCAG 378  
QY 2306 PheArgLeuLeuGluThrGluGluArgThrLysProIleCysValPheTyrAsnHisSer 2325  
DB 379 TTCGGCTCTGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438  
QY 2326 IleLeuValSerGlyThrGlyGlyTyrPserAlaArgGlyCysGluValValPheArgAn 2345  
DB 439 ATCTGTGTAGTGGACAGAGTGGCTGTGGCCAGAGGCTGTGAAGTCTTCCGCAAT 498  
QY 2346 GluSerHisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspVal 2365  
DB 499 GAGAGCACTGACAGCTGCGCAGTGCACCAATGACAGCTTGCCTGTCTCATAGAGTT 558  
QY 2366 SerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThrTyrValIleLeuGly 2385  
DB 559 TCTCGGCGGAGAGATGGGAGAGATCTGCCACCTGAACATGACATACGTGGCTAGGT 618  
QY 2386 ValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArg 2405  
DB 619 GTCACTTGTGGCTGCCCTTGTCTCACCTTCTTCTCATCTCTTGTGGTATCCCTGCCG 678  
QY 2406 SerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuVal 2425  
DB 679 TCCAACCAACAGGAGATCCGACGTAACTGAACAGCTGCCCTGGGCTGAAGCTGGCTC 738  
QY 2426 PheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeu 2445  
DB 739 TTCCTCTGGGATCAACACAGGCTGACCTCCCTTTGGCTGGACAGTCAATGCCATCCG 798  
QY 2446 LeuHisPheLeuThrLeuLysCysThrPheSerThrPheLeuLeuGluAlaLeuHisIleuTyr 2465  
DB 799 GTGACCTTCTGTAACTCTGACACTTTTCTGTGGCTCTGTGAGGCTTGGACCTGTAC 858  
QY 2466 ArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTyrTyrMetLeu 2485  
DB 859 CGGGCACTACCTAGAGTGGCCATGTCAACACGGGCCCATGGCTTCTAATCAATAGCTG 918  
QY 2486 GlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGluGlyTyr 2505  
DB 919 GGCTGGGGGGGTGCTCTTCAATCAAGAGGCTAGCGTGGGCTGGACCCCGAGGGCTTAC 978  
QY 2506 GlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuIleTyrPserPheAlaGly 2525

DB 979 GGAAGCCCTGACTTCTGCTGGCTCTCCATCTATGACAGCTCATCTGAGGTTTCTGGC 1038  
QY 2526 ProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSer 2545  
DB 1039 CGGTGGCCCTTGGCTCTGATGAGTGTCTTCTGTATCATCTGAGCCCGGAGCTTCC 1098  
QY 2546 CysAlaAlaGlnArgGlnGlyPheGluLysGlyProValSerGlyLeuGlnProSer 2565  
DB 1099 TGTGCTGCCAGCGGACAGGGCTTTGAAGAAGATCTGTCTGGGCTGAGACCTTCC 1158  
QY 2566 PheAlaValLeuLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeuLeuSerValAsnSer 2585  
DB 1159 TTGGCCGCTCTCTGCTGCTGAGCGCCAGTGGCTGCTGGACCTGCTCTGTCAACAGC 1218  
QY 2586 AspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnGlyIleGlnGlyProPheIle 2605  
DB 1219 GACACCTCTCTTCTCATCTCTTGTCTATCTGCAATGTGCATCAGAGGCCCTTCAATC 1278  
QY 2606 PheLeuSerTyrValValLeuSerLysGluValArgLysAlaLeuLysLeuAlaCysSer 2625  
DB 1279 TTCCTCTCTATGTGTCTTATGACAGAGAGTCCGGAAGACCTCAAGCTTCCCTGAGC 1338  
QY 2626 ArgLysProSerProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsn 2645  
DB 1339 CGCAAGCCAGCCCTGACCTGTCTGTACCAACCAAGTCCACCTGACCTCGTCTTACAC 1398  
QY 2646 CysProSerProTyrAlaAspGlyValArgLeuTyrGlnProTyrGlyAspSerAlaGlySer 2665  
DB 1399 TGCCCAAGCCCTTACGAGATGGGGGCTGTACCGAGCTTACGGAGACTCGGCGGCTCT 1458  
QY 2666 LeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIleProPheLeuLeuArg 2685  
DB 1459 CTGCACACACCAAGTCCGTGGGAGAGAGTCAAGCCCACTCATCTCTTCTGTGAGG 1518  
QY 2686 GluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeu 2705  
DB 1519 GAGAGTCCGACATGAACTGGCCAGAGGGCCCTGTGCGGGGAGATCAGGCGAGCTG 1578  
QY 2706 PheLeuGluGluGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSer 2725  
DB 1579 TTCCTGAGAGTCAAGACCAAGACATATCTGACAGAGCTCCGACATGATCTGTCTCC 1638  
QY 2726 LeuGluAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGluGlu 2745  
DB 1639 TTAGAAGACGACAGAGTGGCTCTATGCTTACCCCATCATCAGACAGAGAGAGA 1698  
QY 2746 GluGluGluGluGluGluAlaAlaPheProGlyGlyGlnGlyTyrPAspSerLeu 2765  
DB 1699 GAGAGGAG 1758  
QY 2766 GlyProGlyAlaGluArgLeuProLeuHisSerThrProLysAspGlyGlyProGlyPro 2785  
DB 1759 GGGCTGAG 1818  
QY 2786 GlyLysAlaProTyrProGlyAspPheGlyThrThrAlaLysGlySerSerGlyAsnGly 2805  
DB 1819 GGCAGAGGCCCTGTGGCCAGAGACTTGGGACCAAGCAAAAGAGATGGGCAACGG 1878  
QY 2806 AlaProGluGluGluArgLeuArgLysAsnGlyAspAlaLeuSerArgGluGlySerLeuGly 2825  
DB 1879 GCCCTGAG 1938  
QY 2826 ProLeuProGlySerSerAlaGlnProHisLysGlyIleLeuLysLysCysLeuPro 2845  
DB 1939 CCCCTTCAAGGCTCTTCTGCGCCAGCTTCAAAAGGATCTTAAAGAGATGTCTGCCC 1998  
QY 2846 ThrIleSerGluLysSerSerLeuArgLeuProLeuGluGlnCysThrGlySerSer 2865  
DB 1999 ACCATCAAG 2058  
QY 2866 ArgGlySerSerAlaSerGluGlySerArgGlyGlyProProProArgProProProArg 2885

Db 2059 CGGGGCTCCTCCGCTAGTGAGGCGAGCGGGGCGGCCCCCTCCGCCGCCCGCCCGG 2118  
Qy 2886 GlnSerLeuGlnGlnGlnLeuAaenGlyValMetProIleAlaMetSerIleGlyAlaGly 2905  
Db 2119 CAGAGCTCCAGAGCGAGCTGAGCGGGGTCAATGCCCATGCCATGACATCAAGGCGAGC 2178  
Qy 2906 ThrValAspGluAspSerSerGlySer 2914  
Db 2179 ACGGTGATGAGACTGCTCAGGCTCC 2205  
RESULT 13  
ABAI9448  
ID ABAI9448 standard; DNA; 2077 BP.  
XX ABAI9448;  
XX  
XX 23-JUN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 11779.  
XX  
XX Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antispasmodic; antianemic; antiarthritic; cancer;  
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205151.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216860.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239395.  
PR 13-OCT-2000; 2000US-0239397.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 01-DEC-2000; 2000US-0251160.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251899.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-541565/60.  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 PT cancers and metastases -  
 XX  
 PS Disclosure, SEQ ID NO 11779; 1701pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WHO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 2077 BP; 451 A; 586 C; 599 G; 441 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2077  
 Score: 691.00 Matches: 691  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.64% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-916-849a-3 (1-2923) x ABA19448 (1-2077)  
 QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
 DB 3 ACCCAAGTCTTCAAGGTCACAGGCGAGACAGACAGGATGCCGAGAAATGCCCTGCT 62

QY 274 ThrLeuThrIleLeuValThrAspThrAspHisAspProValPheGluGlnGln 293  
 DB 63 ACACTCACCATCTTGATTACTGACACCAATGACCAATGACCTGTGTTCCAGCAGCAGG 122  
 QY 294 TyrIleGluSerLeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAAGAAAGCCACAGGAGAACTGAGAGTGGCTATGAGGTGCTCATCTGCAGGGCC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGly 333  
 DB 183 AGGATGGTAGTCCCTCCCAATGCCAATATCTGTATCCGCTGCTGAGAGGGCTTGG 242  
 QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GGACAGCCCTCTGAAGTCTTTGAGATGACCTGCTGCTGGGGTATCCGACCCGTGCG 302  
 QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
 DB 303 CCGTGGATCGGAAAGAGGTGAATCTTACAGCTGACGGTAGAGGCAAGTACAGAGGT 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsn 393  
 DB 363 CCGGACCCGGGTCTCTCGAGATCCACAGCCGCTGTTTCTTCTTCTGTGAGGATGACAT 422  
 QY 394 AspAsnAlaProGlnPheSerGluValArgTyrValValGlnValArgGluAspValThr 413  
 DB 423 GATTAATGCCCCCAAGTTTATGAGAAAGCGCTATGTGTCTCAGGTGAGGAGATGACT 482  
 QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
 DB 483 CCAGGGGCCCGAGTCTCGAGTCCAGCTCAGCTCGATCCAGACAGAGGAGCAATCCGG 542  
 QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
 DB 543 GTCACATAGCATATGATGATGAGCAATGCTCGGGACAGTTTATCTGATCCCAACT 602  
 QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrLeuArg 473  
 DB 603 GGAGCTCTGAGATGTGTAGGCTCTTGACTATGAGACACCAAGAGTACACCTACGG 662  
 QY 474 ValArgAlaGlnAspGlyValArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
 DB 663 GTGCAGACCAAGAAATGAGGCGGCTCCCACTCTTAATGTCTGCGCTGGTGTGACAGA 722  
 QY 494 GlnValLeuAspIleAspAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
 DB 723 CAGGTCTCGAATATCAAGACCAATGCCCATCTTCTGCAGACACCCCTTCCAGGCTACT 782  
 QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
 DB 783 GTCTGAGAGAGTCCCTTAGAGCTACCTGCTTCCATGTCAGAGCTATGACCGTACT 842  
 QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe 553  
 DB 843 GCTGTGACAAATGCCCGCTGAAATACCGCTTGTGTGGGGTGGACATGACTTCCCTTC 902  
 QY 554 ThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluGluVal 573  
 DB 903 ACCATCAACATGGACACAGGCTGATCTCTGTGCTGCTGAATGACCGGAGAGAAATT 962  
 QY 574 AspPheTyrSerPheGlyValGluAlaArgAspHisGlyTyrProAlaLeuThrAlaSer 593  
 DB 963 GATTTCTACAGCTTGGGGTGAAGCTGAGACCATGACATCCGACATCACTCACTCGCG 1022  
 QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613  
 DB 1023 GCCAGTGTCAAGGTACGTCTCTGATGTCAACGACCAACATTCACCTTAACCAACCA 1082  
 QY 614 GluTyrThrValArgLeuAsnGluAspAlaAlaValGlyTyrSerValValThrValSer 633  
 DB 1083 GAGTATCAAGATGCGGCTCAATGAGGATCACTGTGGGACCAAGGCTGTATCAGGCTGTCA 1142

QY	634	AlaValAspAsgAspAlaHiserVal  lethrTyrgln  lethrSerGlyAsnThrArg	655
Db	1143	GCTGTGACCGGTGATGCTCTATAGTGTATCACTACAGATACCAAGTGGCAATATCTGA	1202
QY	654	AsnArgPheSer  lethrSerGlnserGlyGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	1203	AACCGCTTCTCCATCAACCAAGGAGTGGTGGTGGTGGTATCCCTTGGCCCTGCACCTG	1262
QY	674	AspTyrlValLeuGlnValArgGlnThrValLeuAlaValThrAlaSerAspGlyThrArgGln	693
Db	1263	GACTACAAACTTGAAGCGGAGATGTGTGGCTGTACCGCTCCGAATGGCACTCGGCAG	1322
QY	694	AspThrAlaGln  leuValAlaAsnValThrAspAlaenThrHisArgProValPheGln	713
Db	1223	GACACGGCAGACAGATTGTGTGAATGTCAACGACGCAACCAACCACTGCTCTTTCAG	1382
QY	714	SerSerHisrlyrThrValAlaAsnValAsnGlnAspArgProAlaGlyThrValValLeu	733
Db	1383	AGCTTCCCACTATACAGTAATGTATATAGGACCGGCGGGCAGGCAACCAAGTGGTGGCTG	1442
QY	734	IleSerAlaThrAspGlnValAspThrGlyGlnAsnAlaArg  lethrTyPheMetGlnAsp	753
Db	1443	ATCAGCGCAGAGATGAGACACAGAGTGAAGATGCCGCAATCACTTACTTCAATGAGGAC	1502
QY	754	Ser  leProGlnPheArg  leAspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeu	773
Db	1503	AGCATCCCCCACTGCTCCGATCGATCGATCAACACGGGGGCTGTCAACCAAGCTTAGCTG	1562
QY	774	AspTyrglyAspGlnValSerTyThrIleuAla  leThrAlaArgAspAsnGlyr  lePro	793
Db	1563	GACTATGAAACCAAGATGTCTTACACCTTGCGCATTACTGCTCGGGAACAAATGGCACTTCC	1622
QY	794	GlnIlySerAspThrThrTyrlLeuGln  leuValAlaAspValAlaAspAspAlaPro	813
Db	1623	CAGAAGTCCGACACCACTCACTCGAGATCCGTGGTGAACAGATGAATGACATATGCCCT	1682
QY	814	GlnPheLeuAlaGAspSerTyrglnGlySerValTyrgluAspValProProPheThrSer	833
Db	1683	CAGTTCCTCCAGACACTCTTACCAAGGAGAGTGTCTATGAGATGTGCCACCTTCACTAGC	1742
QY	834	Val  leuGln  leSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyThr	853
Db	1743	GTCTCGACAGATCTCAACCACTGATGTGTATTTCTGACCTTAATGGAGGGTCTTCTACACC	1802
QY	854	PheGlnGlyGlyAspAspGlyAspGlyAspPhe  leValGlnSerThrSerGlyr  leVal	873
Db	1803	TYCCCAAGAGGCGACGATGAGACCGGTACCTTATTTGTGATGTCACCTCAAGGCATCTGTG	1862
QY	874	ArgThrLeuAlaArgLeuAspArgGlnuAsnValAlaGlnTyThrValLeuArgAlaTyThrAla	893
Db	1863	CGAACGCTACGAGAGGCTGATCTGAAGAGAACGTGGCCAGTATGTCTTCCGGGCATATCA	1922
QY	894	ValAspIlyAspGlyMetProProAlaArgThrProMetGlnuAlThrValThrValLeuAsp	913
Db	1923	GTGGCAAGGGGATGCCCCCAGCCCGCACACTTATGAAAGTACAGTCACTGTGTGTGAT	1982
QY	914	ValAlaAspAspAsnProProValPheGlnGlnAspGlnPheAspValPheValGlnGlnAsn	933
Db	1983	GTGAATGACATATCCCCCTGTCTTTTGAAGCAGATGATGTGTGTGTGTGAAGAGAAC	2042
QY	934	SerProIleGlyLeuAlaValAlaArgValThr 944	
Db	2043	AGCCCCATTGGGCTAGCGCTGGCCCCGGGTACACA 2075	
RESULT 14			
AAK83061 standard; DNA; 2077 BP.			
XX	AAK83061;		
XX	AC		
XX	DT		
XX	07-NOV-2001	(first entry)	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37873.		

PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 20-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI, 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 PS Disclosure; SEQ ID NO 37873; 3071bp + Sequence Listing; English.  
 PS  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK7694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM2169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 2077 BP; 451 A; 586 C; 599 G; 441 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2077  
 Score: 691.00 Matches: 691  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.64% Indels: 0  
 DB: Gaps: 0  
 US-09-916-849a-3 (1-2923) x AAK83061 (1-2077)  
 QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
 DB 3 ACCCAAGCTTCAAGGCTCAGCGGACGACCAAGCATGCCCGACGAAGTGCCTGAGCT 62  
 QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlu 293  
 DB 63 ACACTCACCATCTTGTTACGACCAACCATGACCATGACCTGTGTTGACAGCAGCAGAG 122  
 QY 294 TyrTyrGluSerIleuAspGluAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAGAGAGACCTTCAGGAGAACTGAGGTGGCTATGAGGTGCTCAGTCAAGGACC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleuTyrArgLeuLeuGluGlySerGly 333  
 DB 183 ACGAGTGTGATGCCCTCCCAATGCCAATATCTGTACCGCTGCTGAGAGGGGTCTGGG 242  
 QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GGAGACCCCTCTGAAGTCTTTGAGATGACCTCGCTCGGGGGTATCCGAACCGGTGGC 302  
 QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
 DB 303 CCTGTGATTCGGAGAGGTGGAATCTTACCAAGCTGACGAGGAGGACCAAGTACCAAGGCT 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaIleValPheLeuSerValGluAspAspAsn 393  
 DB 363 CGGACCCGGGTCTCGGAGTACCAACAGCCGCTGTTTCTTCTTCTGTGAGAGATGACAT 422



QY	394	AspAsnAlaProGlnPheSerGluValArgGlyValValGlnValArgGluAspValThr	413
Db	423	GATATATGCCCCCACTTATAGTGAAGGCGTATGTGTCCAGGTGAGGAGATGTACT	482
QY	414	ProGlyAlaProValLeuArgValThrAlaSerAspArgAspIleGlySerAsnAlaVal	433
Db	483	CCAGGGGGCCCAAGTACTCCGAGTCCACACTCGATGCCAGCAAGGGAGGAAATCCGTG	542
QY	434	ValHisTyrSerIleMetSerGluValAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr	453
Db	543	GTGACCTATAGCATATAGATAGTGGCAATCTGGGGGACAGTTTATCTGAATGCCAGACT	602
QY	454	GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGlyTyrThrLeuArg	473
Db	603	GGAGCTCTGATATGTGTAGACCTCTTATACATATAGAGCAACCAAGAGTACACCTTAGG	662
QY	474	ValAlaArgAlaGlnAspGlyValArgProProLeuSerAsnValSerGlyLeuValThrVal	493
Db	663	GTGGAGAGCAGAGAGTGGCGGTCCCCACACTTAAATGTCTGGGCTTGGTGACATA	722
QY	494	GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr	513
Db	723	CAGGTCCCTGGAATTCAGCAACATGCCCATCTTCGTACAGACCCCTTTCAGGCTACT	782
QY	514	ValLeuGlnSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp	533
Db	783	GTCCCTGGAGAGTGTCCCTTAGGCTACTCGGTCTCCATGTCCAGGCTATGACGCTGAT	842
QY	534	AlaGlyAspAsnAlaArgLeuGlnTyrArgLeuAlaGlyValGlnValAspPheProPhe	553
Db	843	GCTGTGTACATGGCCCGCTGGAAATACCGCTGTGGGGGTGGACATGACTTCCCTTC	902
QY	554	ThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluGluVal	573
Db	903	ACCATCAACAAATGGCAGCAGGCTGATCTCTGTGGCTGTGAACGTGACCGGAGGAATT	962
QY	574	AspPheTyrSerPheGlyValGlnAlaArgAspHisGlyTyrProAlaLeuThrAlaSer	593
Db	963	GATTTCATACACTTGGGGGTGAAACCTGGAGCAACATGGCACTCCAGACTCATCTGCCTCG	1022
QY	594	AlaSerValSerValThrValLeuAspValAsnAspAsnAspProThrPheThrGlnPro	613
Db	1023	GCCAGTGTCAAGGTGACTGTCTCGGATGTACACGACAAACAACTTTATCCCAACA	1082
QY	614	GluTyrThrValArgLeuAsnGlnAspAlaAlaValGlyThrSerValValThrValSer	633
Db	1083	GAGTACACAGTGCCTCAATAGAGATCAGCTGTGGCACCAAGCGTGTGACGGTGTCA	1142
QY	634	AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGluValAsnThrArg	653
Db	1143	GCTGTGGACCGGATGCTCATATGTGTCAATCACTACAGATACCAAGTGGCAATACTGGA	1202
QY	654	AsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	1203	AACCGCTTCTCATACACCGCAAGCAAGTGGTGGCTGTGTATCCCTTGCCCTGCACATG	1262
QY	674	AspTyrTyrValLeuGlnArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln	693
Db	1263	GACTACAAACTTGACCGGACAGTATGTGGTGTGTACCGCTCCGATGGGCACTCGGAG	1322
QY	694	AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	1323	GACACGGGACAGATTGTGTGTATGTACCCGACCGCAACCCATCGTCTGTCTTTGAG	1382
QY	714	SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu	733
Db	1383	AGCTTCCCACTAAACGTGATATGTTATATAGGACCGCGCGGACGAGCAACGGTGTGTGTG	1442
QY	734	IleSerAlaThrAspGluAspThrArgGluAsnAlaArgIleThrTyrPheMetGluAsp	753
Db	1443	ATCAGCGCCACGGATGAGGACACAGGTGAGAAATGCCCGCATCACTTACTATGAGAGAC	1502
QY	754	SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu	773

Db	1503	AGCATCCCCAGTTCCGCATCCATGACAGACCGGGGCTCTCACCAACCAGCTGAGCTG	1562
Qy	774	AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	1563	GACTATGAAACCAAGAGTCTTACACCTCGGCCATTACTCTGGGGCAATGGCATTTCCC	1622
Qy	794	GlnIlySerSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro	813
Db	1623	CAGAAGTCCACACCACTTACCTGGAGATCTCTGTGACAGACAGTGAATGACATGGCCCT	1682
Qy	814	GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer	833
Db	1683	CAGTTCCTGGAGACTCTTACCGAGGGGAGTGTCTATGAGGATGTGGCAACCTTCATGAC	1742
Qy	834	ValLeuGlnIleSerSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr	853
Db	1743	GTCCCTGAGATCTTCACGACACTGATTCGTGACTTATATGGCAGAGGATCTTCTTACCC	1802
Qy	854	PheGlnGlyValAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal	873
Db	1803	TTCCAGAGAGAGCCACATGAGACGGGATCTTATTTATCTTGATCTCACGCTCAGGATCTGTG	1862
Qy	874	ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla	893
Db	1863	CGAACGCTTACGAGGCTGTGATCCAGAGAACGTGGCCAGATATGTCTTGGCCGGCATATGCA	1922
Qy	894	ValAspArgGlyMetProProAlaArgThrProMetGlnValThrValThrValLeuAsp	913
Db	1923	GTGGACAAAGGGAGATGCCGCCAGCCGACCACTTATGAGATGACAGTCACTGTGTGGAT	1982
Qy	914	ValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGluGluAsn	933
Db	1983	GTGATATGACATCCCTCGTCTTTGAGCAGAGATGAGTTGATGTGTTTGTGAAAGAGAAC	2042
Qy	934	SerProIleGlyLeuAlaValAlaArgValThr	944
Db	2043	AGCCCCATTTGGGCTTACGCCGTGGCCCCGGGTCA	2075
RESULT 15			
AAS31491/c			
ID	AAS31491	standard; DNA; 2077 BP.	
AC	AAS31491;		
XX			
DT			
XX			
XX			
XX	04-DEC-2001	(first entry)	
DE			
XX			
XX	Human DNA for a novel extracellular matrix protein, Seq ID No 570.		
KW	Human; secreted extracellular matrix protein; ds; immunomodulatory;		
KW	Anti-HIV; antihaemic; antirheumatic; anticlerotic; cardiac; vascular;		
KW	cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;		
KW	antialzheimers; immune/autoimmune disease; HIV infection; anaemia;		
KW	human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;		
KW	cancers; hyperproliferative disorder; breast neoplasia; melanoma;		
KW	Serafy syndrome; Gaucher's disease; neurological diseases;		
KW	Alzheimer's disease; Parkinson's disease; cardiovascular disorder;		
KW	cardiac arrest; tachycardia; angina; infection; corneal infections;		
KW	wound healing; immunogen; gene therapy; antisense; food additive.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200155368-A1.		
XX			
PD	02-AUG-2001.		
XX			
XX	17-JAN-2001;	2001WO-US01348.	
PR	31-JAN-2000;	2000US-0179065.	
PR	04-FEB-2000;	2000US-0180628.	
PR	24-FEB-2000;	2000US-0184664.	
PR	02-MAR-2000;	2000US-0186350.	
PR	16-MAR-2000;	2000US-0189874.	



polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotides may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in samples. The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anemias, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). Other uses include wound healing, maintenance of organs before transplantation, support of cell culture of primary tissues, modulation of for example

**Alignment Scores:**

Pred. No.:	0	Length:	2077
Score:	691.00	Matches:	691
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.64%	Indels:	0
DB:	22	Gaps:	0

US-09-916-849A-3 (1-2923) X AAS31491 (1-2077)

Qy	254	ThrHisValPheArgValThrlAglAspHisglMetProArgArgSerAlaLeuAla	273
Db	2075	ACCACAGCTTCAGGGCTACGGCGAGAGACACGGCAGGCCCCACGAGAGGCCCTGGCT	2016
Qy	274	ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlu	293
Db	2015	ACACTCACCACTCTGGTTACTGACACCAATGACATGACCTTGTTCTGACAGCAGAG	1956
Qy	294	TyrIysGlnSerLeuArgIuAsnLeuGluValGlyTyrGluValLeuThrValArgAla	313
Db	1955	TACAAAGAGAGACCTCAGGGAGAACCTGGAGGTTGGCTATGAGGTGCTCACTGCAGGCC	1896
Qy	314	ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrAlaGluLeuGlnGlySerGly	333
Db	1895	ACGATGTGTATGGCCCTCCCAATGCCAATATTCGTACCCCTGCTGAGAGGGGTCTGGG	1836
Qy	334	GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly	353
Db	1835	GGCAGCCCTCTGAACTTTGAGATGCACCTGCTGTGGAGTATCCGAACCGGTGC	1776
Qy	354	ProValAspArgGluGluValGlnSerTyrGlnLeuThrValGluAlaSerAspGlnGly	373
Db	1775	CCTGTGATCGGGAGAGAGTGGATTCACACAGCTGACGCTAGAGGACAACTGACAGGT	1716
Qy	374	ArgAspProGlyProArgSerThrThrAlaValPheLeuSerValGluAspAsn	393
Db	1715	CGGAGCCCGGGTCTCCGGAGTACACAGCCCGCTGTTCTCTTCTGTGAGAGTACAAT	1656
Qy	394	AspAsnAlaProGlnPheSerGlnLysArgTyrValValGlnValArgGluAspValThr	413
Db	1655	GATATATGCCCCCAAGTTTAGTGAGAAAGGCTATGTGCTCCAGGTGAGGAGAAATGTACT	1596
Qy	414	ProGlyAlaProValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaVal	433
Db	1595	CCAAGGGGCCCAAGTACTCCAGTCAACAGCTCGGATCGACGACAGAGGAGCAATCCCTG	1536
Qy	434	ValHisTyrSerIleMetSerGlyAsnAlaArgGlnPheTyrLeuAspAlaGlnThr	453
Db	1535	GTGACATATGACATCATGATGGGAATGCTCGGGGACAGTTTATCTGAGTCCACAGCT	1476

QY	454	G1yAlaLeuAspValValSerProLeuAspTyrGluThrTyrGluTyrThrLeuArg	473
Db	1475	GGAGCTCTGGAGTGGTGGAGCCCTCTTGCATATGAGAGCAACCAAGAGTACACCTTCAGG	1414
QY	474	ValArgAlaGlnAspGlyGlyIArgProProLeuSerAsnValSerGlyLeuValThrVal	493
Db	1415	GTGGAGCACAGGATGATGAGCCGTCGCCCACTCTTAAATGATCTTGGCTTGGTACAGTA	1350
QY	494	GlnValLeuAspTyrLeuAspAsnAlaProIlePheValSerThrProPheGlnAlaThr	513
Db	1355	CAGGTCTCGATATCAACGACATGCCCCCACTTCGTGACACACCCCTTTCAGGCTACT	1280
QY	514	ValLeuGlnSerValProLeuGlyTyrTyrLeuValLeuHisValGlnAlaIleAspAlaAsp	533
Db	1295	GTCTCTGAGAGTGTCCCTTACGCTTACCTGGTCTTCCATGTCACAGGCTATGACGTGAT	1230
QY	534	AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe	555
Db	1235	GCTGGTGAACAATGCCCGCTGGAAATACGCCCTGTGCTGGGGAGACATTCCTCCCTTC	1171
QY	554	ThrIleAsnAsnGlyThrGlyTyrPheSerValAlaAlaGluLeuAspArgGluGluVal	573
Db	1175	ACCATCAACAATGGCACAGGCTGGATCTCTGTGGCTGTCAATGTCACAGGCTATGACGTGAT	1111
QY	574	AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer	593
Db	1115	GATTTCTTAACGCTTTGGGGGTAAAGCTGACAGACATGGCACTTCACGACTCATCTGCTTCG	1050
QY	594	AlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnPro	613
Db	1055	GCCAGTGTACAGGTGACTGTCTCTGATGTCAACGACAACTATCCAACTTATCCCAACCA	996
QY	614	GlnTyrThrValArgLeuAsnGlnAspAlaAlaValGlyThrSerValValThrValSer	633
Db	995	GAGTAAACAAGTGCGGCTCAATAGGATCAAGCTGTGGGACACAGGCTGGTACCGGTCTCA	936
QY	634	AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg	653
Db	935	GCTGTGACCGGATGATCTCATATGTGTCAATCAACGATCAACAAGTGGCAATATCTGA	876
QY	654	AsnArgPheSerIleIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	875	AACCGCTTTCATCAACCAACGCAAGGTGTGGTGGGCTGTATCCCTTGGCCCTGCACTG	816
QY	674	AspTyrGlyLeuGlnIleArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln	693
Db	815	GATTAACAACCTTGAACCGGACAGTATGTGTGGCTGTACCCGCTCGAATGGACCTGGCAG	756
QY	694	AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	755	GACACCGGCAACAATTGGTGAATGTCAACCAACGCAACCCATGCTGCTCTTTCAG	696
QY	714	SerSerHisIleTyrThrValIleAsnValAsnGlnIleAspArgProAlaGlyThrThrValLeu	733
Db	695	AGCTTCCCACTAAACGTGAATGTTTAATAGACCGGCGGAGGACACACGGTGTCTG	636
QY	734	IleSerAlaThrAspGlnAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp	753
Db	635	ATCAGCGCCACCGATGAGGACACAGTGAAGATGCCGCAATCACTACTTATGAGAGAC	576
QY	754	SerIleProGlnPheArgIleAspAlaAspThrGlyValAlaIleThrThrGlnAlaGluLeu	773
Db	575	AGCATCCCCCAAGTCCCGCATGCAATGCAACACGGGGGTGTGCACACCCACGGCTGACCTG	516
QY	774	AspTyrGlnAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	515	GACTATGAAGACCAAGTGTCTTACACCTGTGGCAATTACTGCTCGGACAAATGGCATTC	456
QY	794	GlnIleSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro	813
Db	455	CAGAAATCCGACACCACTTCACTGGAGATCTGTGTGAACGACGTGAATGACAAATGCCCTT	396
QY	814	GlnPheLeuArgAspSerTyrGlnIleSerValTyrGluAspValProProPheThrSer	833

```
Db      395 CAGTTCTTCGAGACTCTCTACCGAGGCAAGTCTTATGAGATGTGCCACCTTCACTAGC 336
Qy      834 VALLLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853
Db      335 GTCTGCAAGATCTCAGCCCACTGATCGTGAATTCGAGCTTAATGGCAGGGTCTTCTACACC 276
Qy      854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal 873
Db      275 TTCCAAGAGGCGCAGCATGAGACGGTGACTTTATTGTGAGTCCACGTCAAGGCATCGTG 216
Qy      874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla 893
Db      215 CGAACGCTTCGAGAGGCTGATCGAAGAAAGTGCCAGTATGCTTGCGGGCATATGCA 156
Qy      894 ValAspLysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913
Db      155 GTGACAAAGGGGATGCCCCAGCCGCCACACCTATGGAAGTGACAGTCACTGTGTGGAT 96
Qy      914 ValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGluGluAsn 933
Db      95 GTGAATGACAAATCCCCCTGCTTTTGAGCAGATGAGTTGATGTGTGTGGAAGAGAAC 36
Qy      934 SerProIleGlyLeuAlaValAlaArgValThr 944
Db      35 AGCCCCATTGGGCTAGCCCGTGGCCCGGGTCAACA 3
```

Search completed: February 14, 2004, 11:59:03  
Job time : 2012 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 11:15:14 ; Search time 302 Seconds

(without alignment)  
4272.063 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923  
Sequence: 1 MRSPTGVPLPPPPPLLL.....AGTDESSGSEPLFFFLH 2923

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
	Ygapop 60.0	Ygapext 60.0
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO/US0916849/runat.11022004.143918.21374/app.query.fasta\_1.3079  
-DB=Issued Patents NA -QFMT=fastep -SUFFIX=Oligo.rnt -MINMATCH=0 -I-LOOPTL=0  
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0916849@cgn 1.169 @runat.11022004.143918.21374 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	0.7	135	1	US-07-998-003A-15
2	21	0.7	135	1	US-08-453-274B-15
3	21	0.7	135	1	US-08-453-695A-15
4	21	0.7	135	1	US-08-453-695A-15
5	21	0.7	135	2	US-08-453-702A-15
6	21	0.7	135	2	US-08-453-702A-15
7	21	0.7	135	5	PCT-US93-12588-15
8	21	0.7	135	5	PCT-US95-08071-15
9	18	0.6	4566	2	US-08-982-412-1
10	18	0.6	4566	2	US-08-982-412-1
11	15	0.5	131	1	US-07-998-003A-82
12	15	0.5	131	1	US-08-453-274B-82

13	15	0.5	131	1	US-08-453-695A-82	Sequence 82, Appl
14	15	0.5	131	1	US-08-268-161A-82	Sequence 82, Appl
15	15	0.5	131	1	US-08-453-702A-82	Sequence 82, Appl
16	15	0.5	131	3	US-09-099-639-82	Sequence 82, Appl
17	15	0.5	131	5	PCT-US93-12588-82	Sequence 82, Appl
18	15	0.5	131	5	PCT-US95-08071-82	Sequence 82, Appl
19	15	0.5	429	3	US-08-905-223-81	Sequence 61, Appl
20	14	0.5	699	3	US-09-276-531-91	Sequence 91, Appl
21	13	0.4	1362	2	US-08-318-837-8	Sequence 8, Appl
22	13	0.4	2023	4	US-09-491-522-6	Sequence 6, Appl
23	13	0.4	2450	4	US-09-491-522-2	Sequence 2, Appl
24	13	0.4	6692	4	US-09-491-522-1	Sequence 1, Appl
25	12	0.4	252	2	US-08-623-906A-1	Sequence 1, Appl
26	12	0.4	1011	4	US-08-811-481-27	Sequence 27, Appl
27	12	0.4	1103	3	US-08-927-219-54	Sequence 54, Appl
28	12	0.4	1946	2	US-08-861-464-1	Sequence 1, Appl
29	12	0.4	1946	2	US-08-396-001-1	Sequence 1, Appl
30	12	0.4	1946	3	US-09-323-433A-1	Sequence 1, Appl
31	12	0.4	2050	4	US-09-620-312D-761	Sequence 761, App
32	12	0.4	2426	3	US-09-028-337-2	Sequence 2, Appl
33	12	0.4	2943	2	US-08-548-159-2	Sequence 2, Appl
34	12	0.4	2994	2	US-08-548-159-4	Sequence 4, Appl
35	12	0.4	3132	2	US-08-224-482-3	Sequence 3, Appl
36	12	0.4	3132	3	US-09-205-921-1	Sequence 1, Appl
37	12	0.4	3132	4	US-09-300-958A-32	Sequence 32, Appl
38	12	0.4	3132	4	US-09-702-705-320	Sequence 320, App
39	12	0.4	3132	4	US-09-736-457-320	Sequence 15, Appl
40	12	0.4	3287	4	US-08-811-481-15	Sequence 5, Appl
41	12	0.4	3311	4	US-09-367-891A-5	Sequence 1, Appl
42	12	0.4	580073	4	US-08-545-528D-31	Sequence 31, Appl
43	11	0.4	36	2	US-08-863-639A-11	Sequence 11, Appl
44	11	0.4	48	4	US-08-589-109A-11	Sequence 3, Appl
45	11	0.4	57	3	US-08-894-511-3	

#### ALIGNMENTS

RESULT 1  
US-07-998-003A-15  
; Sequence 15, Application US/07998003A  
; Patent No. 5643781  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESS: Hicknell  
; STREET: 20 South Clark Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/998,003A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5643781and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 30903  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/346-5750  
; TELEFAX: 312/984-9740  
; TELE: 25-3856  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 135 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-07-998-003A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-07-998-003A-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACTTCCAAAGCTCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 2  
US-08-453-274B-15  
Sequence 15, Application US/08453274B  
Patent No. 5663300

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-274B-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-274B-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACTTCCAAAGCTCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 3  
US-08-453-695A-15  
Sequence 15, Application US/08453695A  
Patent No. 5708143

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,695A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-695A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-695A-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACTTCCAAAGCTCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 4  
US-08-268-161A-15  
Sequence 15, Application US/08268161A  
Patent No. 5798224



GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,161A  
FILING DATE: June 27, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Young J. Sub  
REGISTRATION NUMBER: P-41,337  
REFERENCE/DOCKET NUMBER: 27866/32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-268-161A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
Gaps: 0  
DB: 1

US-09-916-849A-3 (1-2923) x US-08-268-161A-15 (1-135)

QY 483 ProleuseranValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCTCCGCTAGTAACCGTGCAGGCTCTAGACATCAACGACAAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 5  
US-08-453-702A-15  
Sequence 15, Application US/08453702A  
Patent No. 5891706  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-453-702A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
Gaps: 0  
DB: 2

US-09-916-849A-3 (1-2923) x US-08-453-702A-15 (1-135)

QY 483 ProleuseranValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCTCCGCTAGTAACCGTGCAGGCTCTAGACATCAACGACAAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 6  
US-09-099-639-15  
Sequence 15, Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161  
FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 6262237and  
REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-916-849a-3 (1-2923) x US-09-916-639-15 (1-135)

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 3 Gaps: 0

US-09-916-849a-3 (1-2923) x US-09-916-639-15 (1-135)

QY 483 ProLeuserAnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCCGCGTCTAGTAAACCGTCAGGTCCTAGACATCAACGACATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 7  
PCT-US93-12588-15  
Sequence 15, Application PC/TUS9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
PCT-US93-12588-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 5 Gaps: 0

US-09-916-849a-3 (1-2923) x PCT-US93-12588-15 (1-135)

QY 483 ProLeuserAnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCCGCGTCTAGTAAACCGTCAGGTCCTAGACATCAACGACATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 8  
PCT-US95-08071-15  
Sequence 15, Application PC/TUS9508071  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08071  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE: 23 DEC 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-08071-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.72% Indels: 0  
DB: 5 Gaps: 0

US-09-916-849A-3 (1-2923) x PCT-US95-08071-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAla 502  
DB 70 CCACTTCCACGCTTCGGCTAGTACCGTGCAGGCTCCTGACATCAACGACAAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 9  
US-08-465-976A-1  
Sequence 1, Application US/08465976A  
Patent No. 5869632  
GENERAL INFORMATION:  
APPLICANT: SOBET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: ROSEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARBILA, BYRNE, BAIN GILFILLAN, CECCHI  
ADDRESSEE: STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NJ  
COUNTRY: US  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,976A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY P  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 994-1744  
TELEFAX: (201) 994-1700  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 212..2863  
US-08-465-976A-1

Alignment Scores:  
Pred. No.: 3.09e-06 Length: 4566  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.62% Indels: 0  
DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-465-976A-1 (1-4566)

QY 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTirp 2512  
DB 869 GGCCTTGCTGTGGGCTGAGCCCTGAGGCTATGGGAACCTGACTTCTGCTGG 922

RESULT 10  
US-08-982-412-1  
Sequence 1, Application US/08982412  
Patent No. 5958729  
GENERAL INFORMATION:  
APPLICANT: SOBET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: ROSEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE,  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/982,412  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: P181PCT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 212..2863  
US-08-982-412-1

Alignment Scores:  
Pred. No.: 3.09e-06 Length: 4566  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.62% Indels: 0  
DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-982-412-1 (1-4566)

QY 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTirp 2512  
DB 869 GGCCTTGCTGTGGGCTGAGCCCTGAGGCTATGGGAACCTGACTTCTGCTGG 922

RESULT 11  
US-07-998-003A-82  
Sequence 82, Application US/07998003A  
Patent No. 5643781  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSEE: Bicknell  
STREET: 20 South Clark Street

CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/998.003A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5643781and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 30903  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/346-5750  
TELEFAX: 312/984-9740  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-07-998-003A-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-07-998-003A-82 (1-131)

QY 791 G|Y|I|E|P|r|o|g|I|n|y|S|e|r|a|p|h|T|h|T|r|e|n|g|I|n|I|e|u|V|a|l 805  
Db 61 GGCATCCCTCAAAATCAGACACTACTATTGGAAATCTTAGTA 105

RESULT 12  
US-08-453-274B-82  
Sequence 82, Application US/08453274B  
Patent No. 5663300  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
TELEFAX: 312/474-0448  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453.274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-274B-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-274B-82 (1-131)

QY 791 G|Y|I|E|P|r|o|g|I|n|y|S|e|r|a|p|h|T|h|T|r|e|n|g|I|n|I|e|u|V|a|l 805  
Db 61 GGCATCCCTCAAAATCAGACACTACTATTGGAAATCTTAGTA 105

RESULT 13  
US-08-453-695A-82  
Sequence 82, Application US/08453695A  
Patent No. 5708143  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, & Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453.695A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-695A-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0

DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-695A-82 (1-131)

Oy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105

RESULT 14

US-08-268-161A-82

; Sequence 82, Application US/08268161A

; Patent No. 5798224

; GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/268,161A

FILING DATE: June 27, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Young J. Suh

REGISTRATION NUMBER: P-41,337

REFERENCE/DOCKET NUMBER: 27866/32149

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-268-161A-82

Alignment Scores:

Pred. No.: 5.61e-05 Length: 131

Score: 15.00 Matches: 15

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.51% Indels: 0

DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-268-161A-82 (1-131)

Oy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105

RESULT 15

US-08-453-702A-82

; Sequence 82, Application US/08453702A

; Patent No. 5891706

; GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

ADDRESSEE: Borun

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,702A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5891706and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 32657

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-453-702A-82

Alignment Scores:

Pred. No.: 5.61e-05 Length: 131

Score: 15.00 Matches: 15

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.51% Indels: 0

DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-702A-82 (1-131)

Oy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105

Search completed: February 14, 2004, 21:16:08

Job time : 322 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 18:30:55 ; Search time 1943 Seconds  
(without alignment)  
5541.564 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 2923  
Sequence: 1 MRSPTGVPLPPPPPLLL.....AGTVDESSGSEFFLEFLH 2923

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4891994

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-O=/cgn2\_1/USPTO.spool/US09916849/runat.11022004.143919.21449/app.query.fasta\_1.3079  
-DB=Published Applications NA -GMT=fastcap -SUPFIX=Oligo.rmpb -MINMATCH=0.1  
-LOOPT=0 -LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Oligo  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRR=US09916849 @CGN 1.1 1417 @runat.11022004.143919.21449  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBIOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

Result No.	Score	Query Match Length	ID	Description
1	2923	100.0	8772	US-09-788-711A-3
2	2923	100.0	10531	US-10-225-567A-523
3	2837	97.1	8871	US-09-788-711A-1
4	1103	37.7	24370	US-10-292-798-931
5	776	26.5	2332	US-09-764-870-569
6	776	26.5	2332	US-10-125-540-569
7	729	24.9	2391	US-10-176-847-99
8	691	23.6	2077	US-09-764-870-570
9	691	23.6	2077	US-10-125-540-570
10	537	18.4	2603	US-09-843-856-1
11	281	9.6	2695	US-10-264-237-636
12	281	9.6	4152	US-09-925-300-359
13	217	7.4	652	US-09-764-870-571
14	217	7.4	652	US-10-125-540-571
15	131	4.5	509	US-09-764-870-568
16	131	4.5	509	US-10-125-540-568
17	123	4.2	478	US-09-918-995-912
18	118	4.0	658	US-09-764-870-66
19	118	4.0	658	US-10-125-540-66
20	105	3.6	443	US-09-764-870-214
21	105	3.6	443	US-10-125-540-214
22	102	3.5	721	US-09-764-853-931
23	102	3.5	721	US-09-764-853-932
24	102	3.5	721	US-09-764-881-168
25	102	3.5	721	US-09-764-881-169
26	102	3.5	721	US-10-242-747-168
27	102	3.5	721	US-10-242-747-169
28	102	3.5	721	US-10-073-865-152
29	102	3.5	721	US-10-073-865-153
30	102	3.5	754	US-09-764-898-70
31	96	3.3	925	US-09-764-870-70
32	96	3.3	925	US-10-125-540-70
33	82	2.8	290	US-09-764-853-239
34	82	2.8	290	US-09-764-898-142
35	82	2.8	290	US-09-764-881-16
36	82	2.8	290	US-10-242-747-16
37	82	2.8	290	US-10-073-865-49
38	82	2.8	290	US-10-242-747-16
39	79	2.7	11002	US-09-843-856-15
40	70	2.4	321	US-10-017-161-1095
41	61	2.1	444	US-09-843-856-18
42	61	2.1	1856	US-09-843-856-14
43	59	2.0	925	US-10-017-161-1687
44	55	1.9	2201	US-09-764-864-233
45	53	1.8	300	US-09-843-856-19

## ALIGNMENTS

RESULT 1  
US-09-788-711A-3  
; Sequence 3, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Teatra  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788, 711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 8772  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-3

Alignment Scores:  
Pred. No.: 0  
Score: 2923.00  
Percent Similarity: 100.00%

Length: 8772  
Matches: 2923  
Conservative: 0



Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-916-849a-3 (1-2923) x US-09-788-711a-3 (1-8772)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
DB 1 ATGCGAGGCCGGCCACCGGGCTGCCCTCCCAACCGCGCGCGCGCTGCTG 60  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
DB 61 TTGCTGCTCTGCTGCGCGCCGACATTTGGAGAGCAAGTGGGGCTGCTGCTT 120  
QY 41 GlySerArgGlyArgGlySerSerGlyValCysValProMetGlyThrLeuCysProSer 60  
DB 121 GGGTCAGGGAGAGAGGCTTTCGGGGGCTGCGCCCAATGGGCTGCTGCTATCC 180  
QY 61 SerAlaSerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
DB 181 TCAGGCTCGAACCTTGCTCTACACAGCGGCTGCAAGGAGATGCGGGCACTGAGTGA 240  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTTCysProGluSerGlyAlaHis 100  
DB 241 GGCCACTGTGATCCCAACCAAGATGGCTGAGGGTTGGTGTCCAGAAATCCAGGCCCAT 300  
QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
DB 301 ATTCCCTTACACAGAGCTCTGAGAGCTGCTGAGCTGCTGCTGCTGCTGCTGCTG 360  
QY 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGlyIleProCysLeuValAla 140  
DB 361 GGCCACTTTCCTCCCAAGAGGCAAGCTCACACTGCTCCGAGAGAGACCTGTTAAAGGCT 420  
QY 141 ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
DB 421 CCAAGCTCAGAGTCCAGTCTGCAAGCTGGCAAGGCCCCGGGTGAGGGAGGGAGAA 480  
QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgGlyArgAsnValAsnThrAlaProGln 180  
DB 481 AGGTACCAAGAGAGTCCCTGGGTGGCTCGGAAAGAAATGTAAATACAGCCCCCAG 540  
QY 181 PheGlnProProSerGlyGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
DB 541 TTCACAGCCCCCAGCTACAGAGCCAGAGTCCGAGAACCAAGCAGCAGCAGCAGCTGTT 600  
QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGlyAlaGlyArgLeuGlyThrMet 220  
DB 601 GCAATCCCTGAGGGCCATTCACCCGAGCAGAGGTGAGGCAAGTCACTGAGTGAATCA 660  
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
DB 661 GATGGCTCTTTGATAGCCGCTCCACAGTCTTCTCTGAGCCCACTGATCTGATGTA 720  
QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrThrLysSerThrHisValPheArgValThr 260  
DB 721 GTAAACCAAGCCGAGAGCTGATCGTGAAGACCAAGCAACCACTCTTCAGGGCTCAG 780  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 781 GCGCAGAGACAGGAGATGCCGACAGAGGCTTGCTACACTCACTCTTGATGTTACT 840  
QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGlyThrLysGluSerLeuArgGlu 300  
DB 841 GACACCAATGACATGACCTGTGTTCAGAGCAGAGAGTACAGAGAGAGCTCAGGGAG 900  
QY 301 AsnLeuGlnValGlyThrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 901 AACCTGGAGGTGGCTATAGGTGCTCACTGTCAGGCGCAACGATGATGATGCTCTCC 960  
QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGlyGlySerGlyGlySerProSerGlnValPhe 340  
DB 961 AATGCCAATATCTGTATCCGCTGCTGAGGGGTCTGGGGGAGAGCCCTCTGAGAGTCTTT 1020

QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
DB 1021 GAGATGACCTTCGCTCTGGGGTATATCCGAACCGGTGCTCTGTGATCGGGAGAGGTG 1080  
QY 361 GluSerGlyGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
DB 1081 GAATCTTACAGCTGACGATGAGGCAAGTGAACCAAGGGTCCGGAGCCGGGTCTCTGAGT 1140  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
DB 1141 ACCACAGCGCTGTTTCTCTTCTGTGAGGATGACAAATGATATGCCCCAGATTAGT 1200  
QY 401 GlyLysArgValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
DB 1201 GAGAGGCTTATGTGCTCAAGTGAAGAGATGTAATCCAGGGGCCCACTGATCCGA 1260  
QY 421 ValThrAlaSerAspAspArgAspGlySerAsnAlaValIleHisThrSerIleMetSer 440  
DB 1261 GTCAAGCCTCGATCGACAGACAGGGGAGCAATGCCGTGTGCACTATAGCATATGAGT 1320  
QY 441 GlyAsnAlaArgGlyGlnPheThrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
DB 1321 GGCAATGCTCGAGACAGTTTATCTGATGCCAGACTGAGCTGTGATGTGTGAGC 1380  
QY 461 ProLeuAspGlyGluThrThrLysGlyIleThrLeuArgValArgAlaGlnAspGlyGly 480  
DB 1381 CCTGTGACTATGAGACGACCAAGAGTACACCTTACGGGTGCCAGACAGAGATGTGTC 1440  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 1441 CGTCCCCCACTCTAATGTCTGCTGCTGTGTGACAGTACAGTCTCGATATCAACAC 1500  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
DB 1501 AATGCCCATCTTCTGTCAGACCCCTTCCAGGCTATGCTGAGAGGCTCCCTTA 1560  
QY 521 GlyThrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
DB 1561 GGCTACCTGGTTCACAGTCCAGGCTATGACGCTGATGTGAGTCAATGCCGCTG 1620  
QY 541 GluThrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
DB 1621 GAATACCCCTGTGCTGGGTGGACATGACTTCCCTTCAACATCAACAAATGGCAAGGC 1680  
QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheThrSerPheGlyVal 580  
DB 1681 TGGATCTGTGGCTGCTGAATGGAACGGAGAGAAATGATTTCTACAGCTTTGGGTA 1740  
QY 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
DB 1741 GAGCTGAGACATAGGACTCCAGCACTGCTGCGCAGGTGACGCTGACTGTC 1800  
QY 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGlyThrValArgLeuAsn 620  
DB 1801 CTGATGTCAACGACAAATTCACCTTAAACCAACAGATCAACAGTCCGCTCAAT 1860  
QY 621 GluAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis 640  
DB 1861 GAGATGAGCTGTGGGCAACAGCTGTGACGATGCTGTGACCTGTGATCTCAT 1920  
QY 641 SerValIleThrThrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer 660  
DB 1921 AGTGCATCACTTACAGATACCAAGTGGCAATCTCGAAACCCCTTCCATCAACCGC 1980  
QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspGlyThrValLeuGluGln 680  
DB 1981 CAAGTGTGTGTGGCTGTATCCCTTGTGCACTGACATCAACAACTTGAAGGGAG 2040  
QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
DB 2041 TATGTGTTGGCTGTATCCGCTCCGATGGCACTGGGACAGACAGCAATGTGTG 2100

QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrThrValAsn 720  
Db 2101 AATGTCACCGAGCGCAACCCATCGTCTGCTTTCAGAGCTCCACCTATACATGTGAT 2160  
QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuLeuSerAlaThrAspGluAsp 740  
Db 2161 GTTAAATGAGAACCGGCGCGGACGACACGAGTGTGTCTGATACGGGCCACGATAGAGAC 2220  
QY 741 ThrGlyGluAsnAlaArgGlyLeuThrTyrrPheMetGluAspSerGlyProGlnPheArgGlyLe 760  
Db 2221 ACAGGTGAGAAATGCGCGCATCACTTACATGAGAGACAGCATCCCCAGTTCCGCAATC 2280  
QY 761 AspAlaAspThrGluAlaValThrThrGlnAlaGluAspTyrrGluAspGlnAlaSer 780  
Db 2261 GATGCGAGACGCGGGGCTGTCCACCCAGGCTGAGCTGAGATCAAGAAAGCAACATGTCT 2340  
QY 781 TyrrThrLeuAlaLeuThrAlaArgAspAsnGlyLeuProGlnLeuSerAspThrThrTyrr 800  
Db 2341 TACACCTCGGCACTTACTGCTCGGACAAATGGCAATTCGCCAGAGTCCGACACCACTTAC 2400  
QY 801 LeuGluLeuLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrr 820  
Db 2401 CTGGAGATCTCGGTGACGACGTAATGACAAATGCCCTCAGTTCTCGAGACTCTTAC 2460  
QY 821 GlnGlySerValTyrrGluAspValProProPheThrSerValLeuGlnAlaThr 840  
Db 2461 CAGGGCAATGCTTAAAGAGATGTGCACCTTCATCAGCTCGGCAATCTCAGCAATCTCAGCCACT 2520  
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrrThrPheGlnGlyValAspAspGly 860  
Db 2521 GATCGTGAATTCGACCTTAATGCGAGGCTTCTTACACTTCCTCAAGAGGCGGACATGGA 2580  
QY 861 AspGlyAspPheLeuAlaGlnSerThrSerGlyLeuAlaArgThrLeuArgArgLeuAsp 880  
Db 2581 GACCGTGACTTAAATGTTGAGTCCACGTCAAGGCACTGCGCAACGCTACGAGGCTGAT 2640  
QY 881 ArgGluAsnValAlaGlnTyrrValLeuArgAlaTyrrAlaValAspGlyMetProPro 900  
Db 2641 CGAGAGAACGTGGCCGAGATGTCTTGGGAGATATGACAGTGGACAGAGGATGCCCA 2700  
QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspProProAla 920  
Db 2701 GCCCGGCAACCTTAATGAGATGACACTGCTTGGATGTGATGACAAATCCCTCTCTC 2760  
QY 921 PheGlnGluAspGluPheAspValPheValGluGluAsnSerProGlyLeuAlaVal 940  
Db 2761 TTTAGAGCAGATGATGTTGATGTTTGTGAGAGAGAACGCCCAATGGGCTAGCCGTG 2820  
QY 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnLeuMetTyrrGlnLe 960  
Db 2821 GCCCGGTCACAGCCACTGACCCCGATGAGCACCAATGCCCAATATATGTACACAGATT 2880  
QY 961 ValGluGlyAsnLeuProGlyValPheGlnLeuAspLeuPheSerGlyGluLeuThrAla 980  
Db 2881 GTGAGAGGCAACATCCCTGAGGCTCTCCAGCTGACATCTTCTCGGGAGACTGACAGCC 2940  
QY 981 LeuValAspLeuAspTyrrGluAspArgProGlyTyrrValLeuValLeuAlaThrSer 1000  
Db 2941 CTGGTAGACTTACAGTACAGGACCGGCTGATGCTCTGGTCAATCAGGCCACGTC 3000  
QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020  
Db 3001 GCTCCTCTGTGTAGCGCGGCTTACAGTCCAGTCCGCTCTTTCAGCCGAAAGAACCA 3060  
QY 1021 ProValLeuGlyAsnPheGlnLeuPheAsnAspTyrrValThrAspArgSerSer 1040  
Db 3061 CCAATGCTGGGCAACTTGAATCTTTCAACAATGTCTCAACATCGCTCAAGCAGC 3120  
QY 1041 PheProGlyGlyAlaLeuGlyArgValProAlaHisAspProAspLeuSerAspSerLeu 1060  
Db 3121 TTCCCTGGGGGTGCTGATGGCGAGTACCTCCCATGACCTCGATATCTCAGATGACTG 3180  
QY 1061 ThrTyrrSerPheGluArgGlyAsnGluLeuSerLeuValLeuAsnAlaSerThrGly 1080

Db 3181 ACTTAAGCTTTGAGCGGGAATGAATCAAGCTGTGCTTCAATGTCTTCACAGCGGT 3240  
QY 1081 GlnLeuLeuSerSerArgAlaLeuAspAsnAspArgProLeuGluAlaLeuMetSerVal 1100  
Db 3241 GAGCTGAAGCTTACCCCGGCACTGACAAACACCGGCTCTGGAGGCAATAGAGCTG 3300  
QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrLeuLe 1120  
Db 3301 CTGGTGTCAAGCGCGTACACAGCGTACCGCCAGTGGCGGCTGCGGTGATCATTATC 3360  
QY 1121 ThrAspGlnMetLeuThrHisSerLeuThrLeuArgLeuGluAspMetSerProGluArg 1140  
Db 3361 ACCATGAGATGCTCACCCACAGATCAAGCTGGCTGAGAGACATGTCAACCGAGGCG 3420  
QY 1141 PheLeuSerProLeuLeuGlyLeuPheLeuGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
Db 3421 TTCTGTACACCACTGCTAGGCTCTTATTCAGGCGGTGGCCGCACTGCGCACCGCA 3480  
QY 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisLe 1180  
Db 3481 CCGAGCACCGTGGTGTCTTCAAGTACAGGGGAGACCGACGCCCGGGGGGCAATC 3540  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu 1200  
Db 3541 CTCACGCTGAGCTGTGTGGTGGCCAGCCGCAAGGCGCGGGGGCGGCGCTTCTCTG 3600  
QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrrLeuAsnArgSerLeuLeuThrAlaLeuSer 1220  
Db 3601 CCTCTAGAGACTCGAGAGGCGCTTATCTTCAACCGCAAGCTGTGACGCGCATCTCG 3660  
QY 1221 AlaGlnArgValLeuProPheAspAspAsnLeuCysLeuArgGluProCysGluAsnTyrr 1240  
Db 3661 GCACAGCGCGTGGCTCTTCAAGCAACATCTGCTTGGGAGGCTCGAGAGACTAC 3720  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheLeuAlaSerSerSer 1260  
Db 3721 ATGGCTGTGCTGTGGTGGTGGCTGTGACTCTTCGCGCTTCATCAGCGCTCTCTCTC 3780  
QY 1261 ValLeuPheArgProLeuHisProValGlyValLeuArgCysArgCysProProGlyPhe 1280  
Db 3781 GTGCTCTTCGGGCCATTCACCCGCTGAGAGGCTGGCTGCTCCGCTCCGCGCGCTTC 3840  
QY 1281 ThrGlyAspTyrrCysGlnThrGluValAspLeuCysTyrrSerArgProCysGlyProHis 1300  
Db 3841 ACCGGTGACTTACGCGAGACCGAGGTGACCTGTGACTACGCGGCGCTGTGGCCCCAC 3900  
QY 1301 GlyArgCysArgSerArgGluGlyTyrrThrCysLeuCysArgAspGlyTyrrThrGly 1320  
Db 3901 GGGGCTGTGCGAGCGCGAGGCGGCTTACACTGCTCTGTGTATGCTACACAGGCT 3960  
QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly 1340  
Db 3961 GAGCATGTGATGATGATGCTCTCGCTCAGGCGCTTTCAGCCCCGGGTGTCTCAAGAAATGG 4020  
QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
Db 4021 GGCACTGTGTCAACCTGCTGTGGGCGTTCAGATGGCATTTGCGAGACTTC 4080  
QY 1361 GlnLeuProTyrrCysGlnValThrThrArgSerPheProAlaHisSerPheLeuThrPhe 1380  
Db 4081 GAGAGGCTTACCTGACAGGAGTACAGCGGAGCTTCCCGCCCACTCTTCATCACTT 4140  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrTyrrGluArg 1400  
Db 4141 CGGGGCTGTGCGCAGGCTTTCACCTTCACTTCCGCTCTCGTTTGTCAAAAGAGAGCGC 4200  
QY 1401 AspGlyLeuLeuLeuTyrrAsnGlyArgPheAsnGlyHisAspPheValAlaLeuGlu 1420  
Db 4201 GACGGGTGTGTGTGTAACAATGGCGTTCAAAGAGAGATGACTTTGTGGGCGCTTCGAG 4260  
QY 1421 ValLeuGlnGluGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440

Db 4261 GTGATCCAGAGCAGGCTCAGCTCACTTCTCTGACAGGAGGATCAACCAACGATGCTCC 4320  
Qy 1441 ProPheValProGlyValSerAspGlyValTTPH1stHValGlnLeuValTyrTyr 1460  
Db 4321 CCAATGGTGGCCGAGAGATCACTGATGGCACTGGCATACGGTGGAGCTGAAATCTAC 4380  
Qy 1461 AsnLysProLeuLeuGlyValInthGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
Db 4381 AATAAACCACTGTTGGGTGACAGAGGGCTCCCAAGGGCCATCAGAGCAGAGATGGCT 4440  
Qy 1481 ValValThrtValAlaPrgValCyAspThrtGlyValAlaLeuArpHegLysSerValLeuGly 1500  
Db 4441 GTGTGACCGTGTGATGCTGTGACACAGAGTGGCTTGGCTTGGATCTGCTGGGGC 4500  
Qy 1501 AsnTyrSerCysAlaAlaGlnGlyThrtGlnGlyGlySerLysLysSerLeuAspLeuThr 1520  
Db 4501 AACTTATCTCTGCTCTGCTCCAGGGCACTCCAGGGTGGCAGCAGAAAGTCTTGATCTGACG 4560  
Qy 1521 GlyProLeuLeuLeuGlyValProAspLeuProGlyLysPheProValArgMetArg 1540  
Db 4561 GGGCCCTGCTACTAGGCGGGGTGGCTGACCTGCGCCAGAGCTTCCAGTCCGAAATGGCG 4620  
Qy 1541 GlnPheValGlyCysMetArgAspLeuGlnValAspSerThrtAlaLeuMetAlaAsp 1560  
Db 4621 CAGTGTGGTGGCTGACATGGCAAGCTGCAAGTGGACCGGCAATAGCATGGCTGAC 4680  
Qy 1561 PheAlaAlaAsnAsnGlyThrtValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
Db 4681 TTGATTTGCCAACAATGGCACTGCTGCTGGCTGCTCCCAAGAAAGACGTGTGTGACAGC 4740  
Qy 1581 AsnThrtCysH1saAsnGlyValThrtCysValAsnGlnTTPAspAlaPheSerCysGluCys 1600  
Db 4741 AACAATTTGCACATGGGGGCACTGCGTGAACCAAGTGGACCGCTTCACTGAGTGC 4800  
Qy 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlyMetAlaAsnProGlnH1sPheLeu 1620  
Db 4801 CCCCCTGGCTTTGGGGCAAGCTGGCCCAAGAAATGGCAATTCACAGCACTTCTCG 4860  
Qy 1621 GlySerSerLeuValAlaTTPH1sgLysSerLeuProLysSerGlnProTTPYLeu 1640  
Db 4861 GGCAGCAGCTGGTGGCTGGCATGGCTCTGCTGGCCATCTCCCACTGGTGTACTC 4920  
Qy 1641 SerLeuMetPheArpThrtArgGlnAlaAspGlyValLeuLeuGlnAlaAlaThrtArgGly 1660  
Db 4921 AGCTTCAATGTTCCGACAGCCAGCCAGCCAGCGGTCTGCTGCAAGGCATCACAGGGGG 4980  
Qy 1661 ArgSerThrtLethrLeuGlnLeuArgGlyValH1sValMetLeuSerValGlyGlyThr 1680  
Db 4981 CGCAGCACCATCACTCAAGCTHCGAAGGGCCACGTGATGCTGACGCTGGAGGGCACA 5040  
Qy 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyArgAlaAsnAspGlyAspTTPH1s 1700  
Db 5041 GGGCTTCAAGGCTCTCTCTCTCGCTGGAGCCAGGCCAGGCCCAATGACGTGACTGGCAC 5100  
Qy 1701 H1sAlaGlnLeuAlaLeuGlyAlaSerGlyValProGlyH1sAlaAlaLeuSerPheAsp 1720  
Db 5101 CAGTGCACAGTGGACATCGGAGGCCAGGGGGGGCTGGCCATGCTGCTCTTCTGAT 5160  
Qy 1721 TyrGlyValGlnArgAlaGlyValAsnLeuGlyProArpLeuH1sgLysLeuH1sLeuSer 1740  
Db 5161 TATGGCAGCAGAGAGAGAGAGGGGCACTGGGGCCCGGCTGCAATGATCTGCACTGAGC 5220  
Qy 1741 AsnLethrtValGlyGlyLeuProGlyProAlaGlyValAlaAlaArgLysPheArgGly 1760  
Db 5221 AACAATAACATGGGGCGGAATACTGGGGCCAGCCGGCTGTGGCCCTTTCGGGGC 5280  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrtProGlnGlyValAsnSerLeuAspProSer 1780  
Db 5281 TGTGTGAGGGGTGGTGGGTGAGGATACCGCAGAGGGGGGTTPAACACCTCGAGTCCAGC 5340  
Qy 1781 H1sgLysLysSerLysAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGAGACATCAAGTGGAGCAAGGCTGTAGCTGCTGACCTTGTGACTCAAC 5400

Qy 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTTPAspSerTyrSerCysSerCysAsp 1820  
Db 5401 CCGTGTCCGCTTACAGCTATTTGACAGACAGACTGGAGACGCTATCTTCGACGCTGTAT 5460  
Qy 1821 ProGlyTyrTyrGlyAspAsnCysThrtAsnValCysAspLeuAsnProCysGlyH1sgL1n 1840  
Db 5461 CAGGTTACTATGATGACAACTGTATGTGTGTGACTGAACCCGTTGATGAGCCAG 5520  
Qy 1841 SerValCysThrtArgLysProSerAlaProH1sgLysTyrThrtCysGlyLysProProAsn 1860  
Db 5521 TCTGTGTACCCGCAAGGCCAGTGGCCCCCAATGGCTATACCTGGCAGTGTCCCCCAAT 5580  
Qy 1861 TyrLeuGlyProTyrCysGlyLysThrtArgLysAspGlnProCysProArgGlyTTPTPGly 1880  
Db 5581 TACCTTGGCCATTACTGTAGACAGAGANTTACAGACCTTGTCCCGTGGCTGGTGGGA 5640  
Qy 1881 H1sProThrtCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn 1900  
Db 5641 CATCCCACATGTGGCCCATGTGCAACTGTGATGTCAAGCAAGGCTTGTGACCAAGCTGCAAC 5700  
Qy 1901 LysThrtSerGlyGlyCysH1sCysLysGlyAsnH1sTyrArgProProGlyLysSerProThr 1920  
Db 5701 AAGACAGCGCGAGGTGCATGCAAGAGAAACACTACCGGCCCCAGGCAAGCCACCC 5760  
Qy 1921 CysLeuLeuCysAspCysTyrProThrtGlySerLeuSerArgValCysAspProGlyAsp 1940  
Db 5761 TGCCTCTTGTGTGACGTGCTACCCCAAGGCTCTCTTGTCAAGATGTGTGACCTGAGAT 5820  
Qy 1941 GlyLysCysProCysLysProGlyValL1sgLysArgGlnCysAspArgCysAspAsnPro 1960  
Db 5821 GGCAGGTGTCCATGGAAGCAGAGTGTCAATGGGCGCTCAGTGTGACCGCGTGTACAACCTT 5880  
Qy 1961 PheAlaGlnValThrtThrtAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaL1e 1980  
Db 5881 TTTCGTGAGGTACCAACCAATGCTGTGAAAGTAAATTAATGACAGCTGCCCAAGAGGAT 5940  
Qy 1981 GlnAlaGlyL1eTTPProArgThrtArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
Db 5941 GAGGCTGGGATCGTGGGCCCCGATCCGCTTCGGGCTGCTGCTGCTGCTGCTGCTGCTG 6000  
Qy 2001 LysGlySerPheGlyThrtAlaValArgH1sCysAspGlyH1sArgGlyTTPLeuProPro 2020  
Db 6001 AAGGCTCTTTGGAGATGCTGTGGCCACTGTGATGAGCAAGAGGGGGTGGCTCCCCCA 6060  
Qy 2021 AsnLeuPheAsnCysThrtSerLethrtPheSerGlyLeuLysGlyPheAlaGlnArgLeu 2040  
Db 6061 AACCTTTCACTGACAGTGCATCACTTCTCAAGAACTGAAGGGCTTGGCTGAGGGCTTA 6120  
Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg 2060  
Db 6121 CAGCGGAATGAGTCAAGGCTTCACTCAAGGCGCTCCCAAGCACTAGCCCTGCTCCGCC 6180  
Qy 2061 AsnAlaThrtGlnH1sThrtAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
Db 6181 AAGGCCACGAGACACAGCTGCTTACCTTGGAGAGAGTCAAGGTGGCTTACCAAGCTG 6240  
Qy 2081 AlaThrtArgLeuLeuAlaH1sGlySerThrtGlnArgGlyPheGlyLeuSerAlaThrtGln 2100  
Db 6241 GCCAGCGGCTGTGGCCCAAGAGACCAAGCGGGGCTTTGGGCTGTCTGCAACAG 6300  
Qy 2101 AspValH1sPheThrtGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrtAlaAsn 2120  
Db 6301 GAGGTGCACCTTCACTGAGATCTGCTGGCGGTGGGAGGCCCTCTGTGACACAGCAAC 6360  
Qy 2121 LysArgH1sTTPGlyLeuL1eGlnGlnThrtGlnGlyGlyTTPH1sAlaThrtLeuLeuGlnH1s 2140  
Db 6361 AAGCGCACTGGAGGTGATTCAGCAGAGAGAGGGTGGCACTGGCTGGCTCCAGCAC 6420  
Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgH1sThrtTyrLeuSerProPhe 2160  
Db 6421 TATGAGGCTTACGCGAGTCCCTGGGCCAGAACTGGGCAACCTTACCTTAAGCCCTTTC 6480

2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspIleGlyAsnPheAla 2180  
Db ACCATGTCACCCCAACATTGTCACTTCGTAAGCGCTTGACAAAGGAACTTGTGCT 6540  
Qy 2181 GlyAlaIleLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAsnIleGluThr 2200  
Db GGGGCGCAAGCTCCCGCTACAGAGCCCTGCGTGGAGACAGCCCGGACCTTGAGACA 6600  
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
Db ACAGTCATCTCTCGTAGAGTCTGTCTTCAAGAGAGAGCCCGGTGGTCAAGGCCCGAGGC 6660  
Qy 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgGlnArgArgHisProGluLeu 2240  
Db CCGGAGAGGCGCCAGAGAGCAAGAGAGCTGGACCGGACAGCGAGCCCGAGAGCTG 6720  
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db AGCCAGGCTGAGGCTGTGGCCAGCTCATCTTACCGACCCCTGGCCGAGCTACTGCT 6780  
Qy 2261 HisAsnTyrAspProAspIleArgSerLeuArgValProIleArgProIleIleAsnThr 2280  
Db CATACTATGACCTGACAGAGCGAGCTTGAGAGTCCCAAGCCCGATCATCAACACA 6840  
Qy 2281 ProValIleSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspIle 2300  
Db CCGGAGGAGAGCATGCGCTCATGATGATGAGAGCTTCTGCCCCGGGCGCTTGACAA 6900  
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleGlyVal 2320  
Db CCGGTCACGGTCAATTCGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6960  
Qy 2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrSerAlaArgIleCysGlu 2340  
Db TTCTGGAACCAATTCATCTGTGTGTCAGTGGCACAGGTGGTGGTGGCCAGAGCTGTA 7020  
Qy 2341 ValValIlePheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
Db GTGCGCTTCCCGCAAG 7080  
Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuIleThrLeuThr 2380  
Db GTGCTCATGAG 7140  
Qy 2381 TyrValAlaLeuGlyValThrLeuAlaIleLeuLeuThrPhePheLeuThrLeu 2400  
Db TACGGGCTCTAGAGGTCACTTGGCTGCTTGTCTGCTCACTTCTTCTTCTCACTCTC 7200  
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleGly 2420  
Db TTGCGTATCCCTGGCTCCACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7260  
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
Db CTGGCTCAGCTGTCTTCTCTCTGAGAAATCAAGAGCTGACCTCTTGTGCTGACA 7320  
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTyrAlaLeuLeuGlu 2460  
Db GTCATTTGCCATCTGCTGCACTTCTGTACCTCTGCACTTCTTCTGAGGCTGTGTAAG 7380  
Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
Db GCGTTGACCTGTATCCGGGACATCTAGAGTGCAGATGTCACACCGGCGCCCATGTGGC 7440  
Qy 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db TTCTACTACAGCTGGGCTGGGAGCTGTGCTTCTTCAACAGGCTTACGCTGGAGCTG 7500  
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTyrLeuSerIleTyrAspThrLeuIle 2520  
Db GACCCGAGGAGGCTACGGGAGACCTGACTTGTGCTGCTCATCATGACAGCGTCAATC 7560  
Qy 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540

7561 TGAAGTTTGTGGGCGCCGGTGGCTTTGCGCTTCGAGAGAGTCTTCTGTACATCTCG 7620  
Qy 2541 AlaAlaArgAlaSerCysValAlaGlnArgGlnGlyPheGlyValArgGlyProAlaSer 2560  
Db GGGGCCGGGCGCTCTGTGTGCTGCCAGCGGAGGAGGCTTGTAGAAAGATCTCTCTCG 7680  
Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeu 2580  
Db GGCGCTGAGAGCCCTCTTGTGCGGTCTCTGTGTGTGAGCGGACAGTGGCTGTGACCTG 7740  
Qy 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrIleLeuPheAlaThrCysAsnCysIle 2600  
Db CTCTCTGTCAACAGAGACACCTCTCTTCCATCACTCTTGTGTACCTGTACATTTGACATC 7800  
Qy 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGlyValArgIleValAlaLeu 2620  
Db CAGGCGCCCTTCAATCTTCTCTCTTATGTGTGTGTACCAAGAGAGTCCGAAAGCACTC 7860  
Qy 2621 LysLeuAlaCysSerArgIlePProSerProAspProAlaLeuThrThrIleSerThrLeu 2640  
Db AAGCTTGTCTGACCGCGAAGCCCAAGCCCTGACCTGTCTGACCAACAGTCCACCTGTG 7920  
Qy 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db ACCTGTCTCTCAACTGTGCCAGCGCCCTTACAGAGATGGCGGCTGTACCAAGCCCTTACGGA 7980  
Qy 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
Db GACTCGCGGAGCTCTCTGACAGAGACAGTGTCTGGGCAAGAGAGAGAGAGAGAGAGAG 8040  
Qy 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db CCTTCTTGTGTGAGAGAGAGAGTCCGACATGAACTTGGCCAAAGGCGCCCTGGCTGGG 8100  
Qy 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db GATCAGAGAGCTGTCTCTGGAAGTCAAGACAGAGAGATGATCTCGACAGAGAGAGAG 8160  
Qy 2721 AspSerAspLeuSerLeuGluAspAspGlnSerIleSerTyrAlaSerThrHisSerSer 2740  
Db GACAGTACCTGTCTTAAAG 8220  
Qy 2741 AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
Db GACAGTGTGAG 8280  
Qy 2761 TrpAspSerLeuLeuGlyProGlyValArgIleArgLeuProLeuHisSerThrProLysAsp 2780  
Db TGGAGATGCTGTGTGGGCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 8340  
Qy 2781 GlyGlyProGlyProGlyLysAlaProTyrProGlyLysAspPheGlyThrThrAlaGlyGlu 2800  
Db GGGGCGCCAGGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8400  
Qy 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
Db AGTAGTGGCAACGGGAGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8460  
Qy 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisGlyGlyIleLeuLys 2840  
Db GAGGAGTCCCTAG 8500  
Qy 2841 LysLysCysLeuProThrIleSerGlyLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
Db AAGAGATGTCTGCGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8560  
Qy 2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
Db TGCAAGAGAGTCTTCCGGGAGCTCTCCGCTAGTAGAGAGAGAGAGAGAGAGAGAGAGAG 8640  
Qy 2881 ArgProProProAlaGlnSerLeuGlnGluLeuAsnGlyValMetProIleAlaMet 2900

Db 8641 CGCCACCGCCGCGAGAGCTCCAGAGACGTGACAGCGGCTCATGCCCATCGCCATG 8700  
Qy 2901 SerIleValaIagIyThrValaAapSerSergIyubheuphean 2920  
Db 8701 AGCATCAAGGACGACGAGTGAAGAGACTCGTCCAGCTCCGAAATTCCTCTTAAAC 8760  
Qy 2921 PheLeuHis 2923  
Db 8761 TTCCTCAT 8769

RESULT 2  
US-10-225-567A-523  
; Sequence 523, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:  
; APPLICANT: LifeSpan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burmer, Glena C.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 523  
; LENGTH: 10531  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-225-567A-523

Alignment Scores:  
Pred. No.: 0 Length: 10531  
Score: 2923.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-916-849A-3 (1-2923) x US-10-225-567A-523 (1-10531)

Qy 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
Db 63 ATCGGAGCGCGGCGACCGGCGTCCCTCCCAACGCGCGCGCGCTGCTGCTG 122  
Qy 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyValAspGlnValGlyProCysArgSerLeu 40  
Db 123 TTGCTGCTGCTGCTGCGCGGCACTATTGGAGAACCAAGTGGGGCCCTGCTTCCTTG 182  
Qy 41 GlySerArgGlyArgGlySerSergIyAlaCysAlaProMetGlyThrPheLeuCysProSer 60  
Db 183 GGGTCCAGGGGAGCAAGGCTTTCGGGGGCTCTGGCCCCCATGGGCTGCTCTTCATCC 242  
Qy 61 SerAlaSerAlaLeuThrLeuLeuYThrSerArgCysArgAlaAlaGlyThrGluLeuThr 80  
Db 243 TCAGCGCTGAACCTTGCTGCTACACCAAGCGCTGCGAGATGGGGCACTGAGCTGACT 302  
Qy 81 GlyHisLeuValProHisHisAspGlyLeuArgValTProCysProGluSerGluHis 100  
Db 303 GGGCACTGGTACCCCAACGAGTGGCTAGAGGTTGGTCCAGAAATCCAGAGCCCAT 362  
Qy 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
Db 363 ATTCCCTTACCAACAGCTCTCTGAAGCTGCCCCCTGAGAGCTGCTCCCTCGGGCATTTGA 422  
Qy 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGluGluHisProCysLeuValAla 140  
Db 423 GGGCACTTTTCCCAACGAGGAGCTCACTGCCCAAGAGACCCGCTTAAAGGCT 482  
Qy 141 ProArgLeuArgCysGlnSerCysValLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160

Db 483 CCAAGGCTCAGATGACAGTCTCTGCAAGCTGACAGAGCCCGGCGCTCAGAGGAGGGA 542  
Qy 161 ArgSerProGluGluSerLeuGlyValArgArgValAsnThrAlaProGln 180  
Db 543 AGTCACCAAGAAAGTCTCCGAGTGGCGTCCGAAAGAAAGTAAATACACCCCCCAG 602  
Qy 181 PheGlnProProSerSergIyGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
Db 603 TTCAGACCCCAAGCTTACCAAGGACCAAGTCCGAGAACCAAGCCAGAGGACCCCTGTT 662  
Qy 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyValLeuGluIleuThrMet 220  
Db 663 GCATCCCTGAGGGCCATGACCCGACGAGGTGAGGAGGTGACTGAGATGACCAAG 722  
Qy 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
Db 723 GATGCCCTTTGATAGCCGCTCCAAACGATTCCTCCCTGAGACCCAGTCACTGATGA 782  
Qy 241 ValThrThrAlaGluGluLeuAspArgIleThrIleSerThrHisValPheArgValThr 260  
Db 783 GTAACCAACAGCGAGAGGTGATGCTGAGACCAAGACCAACCACTTCAGGGTCAAG 842  
Qy 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleValThr 280  
Db 843 GCGCAGAGACCAAGGATGCCCGACGAAAGTGCCTGCTACACTCACCATTTGGTTACT 902  
Qy 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluIleuThrValLeuArgGlu 300  
Db 903 GACACCAATGACCAAGACCTGCTGTGTTCAGACAGAGATACAGAGAGCTTCAGGGAG 962  
Qy 301 AsnLeuGluValaGlyValGluValLeuThrValaArgAlaThrAspGlyAspAlaProPro 320  
Db 963 AACCTGAGAGTTGGCTATAGAGGTGCTCACTGTCAGGGCCAGAGATGTATCCCTCC 1022  
Qy 321 AsnAlaAsnIleLeuThrArgLeuLeuGluGlySerGlySerProSerGluValPhe 340  
Db 1023 AATGCCAATATTCGTACCGCTGCTGAGAGGGGTCTGGGGGAGGCCCTCTGAAGCTTT 1082  
Qy 341 GluIleAspProArgSergIyValIleArgThrArgGlyProValAspArgGluGluVal 360  
Db 1083 GAGATCGAACCTCGCTGCGGGTGTATCCGAACCCCTGCGCTGTGATCGGGAAGAGTG 1142  
Qy 361 GluSerYThrGlnLeuThrValaGluAlaSerAspGlnIleArgAspProGlyProArgSer 380  
Db 1143 GAATCCTACAGCTACGCTAGAGGCAAGTGAACAGGGTCCGGGATCCCGGATCTCGAGT 1202  
Qy 381 ThrThrAlaAlaValPheLeuSerValaGluAspAsnAspAsnAlaProGlnPheSer 400  
Db 1203 ACCACAGCGCGTGTTCCTTCTGTGAGAGATGAACAATGATATATGCCCCCAAGTTAAT 1262  
Qy 401 GluIleArgIleValaGlnValaArgGluAspValThrProGlyAlaProValLeuArg 420  
Db 1263 GAGAGCGGTATGTGTCAAGTGAAGAGATGTAGTCCAGGGGCCCAATATCTCCGA 1322  
Qy 421 ValThrAlaSerAspArgAspIleGlySerAsnAlaValaHisIleIleSerIleMetSer 440  
Db 1323 GTACACGCTCCGATCGAGACCAAGGGGAGCAATGCGTGTGTCACATATGACTATGAGT 1382  
Qy 441 GlyAsnAlaArgGlyGlnPheThrLeuAspAlaGlnThrGlyAlaLeuAspValaIleSer 460  
Db 1383 GGCATGCTGGGAGACAGTTTATCTGATGAGGCCCAAGCTCTGATGTGTGATGAC 1442  
Qy 461 ProLeuAspIleGluThrThrIleArgIleuThrLeuArgValaArgAlaGlnAspGly 480  
Db 1443 CCTCTTGAATAGAGAGCAAGAGATGACCTTACCGGGTCCAGACCAAGATGTGTGC 1502  
Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1503 CGTCCCACTCTCTTAATGTCTGTGCTTGAGAGATGACAGGTCTGTGATCAAGCAGAC 1562  
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
Db 1563 AATGCCCAATCTGTGTACGACACCTTTCCAGGCTACTGTCTGTGAGAGCGTCCCTTA 1622



QY	521	GLYTYRLeuValIleuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540	QY	881	ArgGluAsnValAlaGlnTyRValLeuArgAlaTyRAlaValAspLysGlyMetPro	900
Db	1623	GGCTACTGCTGTTCTCATGTCCAGGCTATCGACGCTGATGCTGTGACATGCGCCCTG	1682	Db	2703	CGAGAGAAAGTGGCCAGATATGTTGCGGGCATATGACAGTGCACAAAGGAGATGCCCA	2762
QY	541	GLUTYRArgLeuAlaGlyValGlyHisAspPheProPheThrTleAsnGlnGlyThrGly	560	QY	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal	920
Db	1683	GAATACCGCTGCTGGGGGGGACATGACTTCCCTTCAACATCAAGATGGCACAGGC	1742	Db	2763	GCCCGCACACCTATGAAAGTGAAGTCACTGTGGTGGATGTGAATGACATATCCCCCTGTC	2822
QY	561	TPPLeuSerValAlaAlaGluLeuAspArgGluGluValAspPheTySerPheGlyVal	580	QY	921	PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal	940
Db	1743	TGGATCTGTGGCTGCTGACTGACCGGAGAGAGTTGATTCTACAGCTTTGGGGTA	1802	Db	2823	TTTGAAGAGATGACTTTGATGTGTGTGGAAGAGAACAGCCCATTTGGGCTAGCCGTA	2882
QY	581	GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600	QY	941	AlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyRLeuIle	960
Db	1803	GAAGCTCGAGACCATGCGCACTCGACGACTCGCTCGCGCCAGTCTCAGCGGATGTGTC	1862	Db	2883	GCCCGGTCACAGCCACTGACCCCGATGAGGACCAATGCCCATATATGACAGATT	2942
QY	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyThrValArgLeuAsn	620	QY	961	ValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla	980
Db	1863	CTGGATGTCAAGACAAACATCCACTTTACCAACAGATGACAGTGGCGGCTCAAT	1922	Db	2943	GTTGAGGGGACATCCCTGAGGTCCTTCAGCTGACATCTTCCGGGGAGCTGACAGCC	3002
QY	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640	QY	981	LeuValAspLeuAspTyRValAspArgProGluTyRValLeuValIleGlnAlaThrSer	1000
Db	1923	GAGGATGCGAGCTGTGGGACACAGCGGTGAGCGGTGCAAGTGTGACCGTATGTCAAT	1982	Db	3003	CTGGTACCTTGAAGCTAAGAGGACCGGCTGATACGTCTGTGATCCAGGCCACGTCA	3062
QY	641	SerValIleThrTyRAlaIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660	QY	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	1983	AGTGTGATCACTTACCAAGATCAACAGTGGCAATACTGAAACCGCTTCTTCATCAACAGC	2042	Db	3063	GCTCCTCTGTGAGCCGGGCTACAGTCCAGCTCCGCTCTTGAACCGCAATGACAAACCA	3122
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyRtyLeuGlnIleArgGln	680	QY	1021	ProValLeuGlyAsnIlePheGluIleLeuPheAsnThrValThrAsnArgSerSer	1040
Db	2043	CAAAAGTGTGGTGGGCTGGATTCCTTGGCTGGCCACTGACATCAAACTTGAACGGGCAAG	2102	Db	3123	CGAGTGTGGGACATTTGAGATCTTTCACAACTATGTGACAAATGCTCAAGGACGC	3182
QY	681	TyRValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700	QY	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	2103	TATGTGTTGGCTGTTAACCGCTCCGATGGCACTCCGCAAGACACGCAAGATTGGGTG	2162	Db	3183	TTCCCTGGGGGTGCCATTGGCGAGATCACTGCCATGACCTGATATCTCAGATATGCTG	3242
QY	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyRThrValAsn	720	QY	1061	ThrTyRSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080
Db	2163	AATGTCAACCGAGCCCAACCACTCCTCTTCAGAGCTCCCACTATACAGTCAAGTAAAT	2222	Db	3243	ACTTACAGCTTTGAGCGGGGAAATGAACTCAGCCGTGGTCTGCAATAGCTCCACGGGT	3302
QY	721	ValAsnGluAspArgProAlaGlyThrThrValIleValIleSerAlaThrAspGluAsp	740	QY	1081	GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	2223	GTTAATGAGGACCGGCGGCGGACAGGACACAGTGTGCTATGACCGCCACGAGATGGAGAC	2282	Db	3303	GAGCTGAGCTTAAGCCGGCGACTGGACAAACACGGCTCTGGAGGCACTATGAGGCTG	3362
QY	741	ThrGlyGluAsnAlaArgIleThrTyRPhMetGluAspSerIleProGlnPheArgIle	760	QY	1101	LeuValSerAspGlyValAlaHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Db	2283	ACAGGTGAGAAATGCCCGCATCACTTCAATGAGGACAGCATCCCCAGTTCCGCGATC	2342	Db	3363	CTGGTGTCAACGGGTGTACACAGGTGACCGCCAGTGGCGGTGCGTGTACATCATCTC	3422
QY	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyRValAspGlnValSer	780	QY	1121	ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg	1140
Db	2343	GATGACAGACCGGGGGCTGTCAACCCAGGCTGAGCTGACATCAACAAACCAAGTGTCT	2402	Db	3423	ACCGATGAGATGCTCACCCACAGATCACTGCGGCTGGAGAGCATGTCAACCCGAGCGC	3482
QY	781	TyRThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerSerThrThrTyR	800	QY	1141	PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Db	2403	TACACCTTGGCCATTAATCTCTGGGACAAATGGATATCCCAAGAGTCCGACACACTTAC	2462	Db	3483	TTCTGTACCACTGTCTAGGCTCTTCAATCCAGGGGGTGGCGGCACTGACACGCA	3542
QY	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyR	820	QY	1161	ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle	1180
Db	2463	CTGGAGATCTGTGTGACACAGCGTGAATGACATGGCCCTCAAGTCTCTGGAGAACCTCTAC	2522	Db	3543	CCGAGCACAGTGTGTCTTCAAGCTACAGGGGACACCGACGCCCGGGGGGCAACATC	3602
QY	821	GlnGlySerValTyRValAspValProProPheThrSerValLeuGlnIleSerAlaThr	840	QY	1181	LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu	1200
Db	2523	CAGGGGACGTGTATGAGATGTGCCACCTTCACTAGCGTCTGTGAGATCTCACGCACT	2582	Db	3603	CTCAACGTGACCTGTGTGTGGGCGAGCGCCAGGGGCGGGGGGCGCCCTTCTCTG	3662
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyRThrPheGlnGlyGlyAspAspGly	860	QY	1201	ProSerGluAspLeuGlnGluArgLeuTyRLeuAsnArgSerLeuLeuThrAlaIleSer	1220
Db	2583	GATCTGATCTTGGACTTAATGGCAGGATCTTCTACACTTCCAAAGAGGCAACATGGA	2642	Db	3663	CCCTTGAAGACCTGCGAGAGCGCTTATCTTCAACCGCACTGTGCTGACGGCATCTCG	3722
QY	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880	QY	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGlyAsnTyR	1240
Db	2643	GACGGGTGACTTTATGTTGATGTCAAGTCAAGGACATGTGGAACGCTGAGGGGTGGAT	2702	Db	3723	GCAAGGCGGTGTGCTGCTTGCACGACAACTGTGCTGGGGAGGCCCTGGAGAGACTAC	3782
QY	2643	GACGGGTGACTTTATGTTGATGTCAAGTCAAGGACATGTGGAACGCTGAGGGGTGGAT	2702	QY	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260



Db	3783	ATGGGCTGCGGTGGCTGCTGCGCTTCGACTTCCTCCGCGCTTCCTCANTGACTCTCTCC	3842
QY	1261	ValLeuPheArgProIleHisProValGlyLeuArgCysArgCysProProGlyIle	1280
Db	3843	GTGCTCTCCGGCCCATCAACCCGGTGGAGGGCTGGCTGCGCTGCGCCGGCTTC	3902
QY	1281	ThrGlyAspArgIleCysGluThrGluValAspLeuGlyTyrSerArgProCysGlyProHis	1300
Db	3903	ACGGGTGACTACTGGAGACCGAGGTGACCTCTGCACTCGAGCGCTTGTGGCCCCAC	3962
QY	1301	GlyArgCysArgSerArgGlyGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly	1320
Db	3963	GGGGGCTGCGCGACGCGCGAGGGGGCGGTCAACTGCTCTCTGTAAGGCTTACGGGT	4022
QY	1321	GluHisCysGlyValSerIleArgSerGlyArgCysThrProGlyValCysGlyAsnGly	1340
Db	4023	GAGCACTGTGAGGTGAGTCTGCTGCACGAGCGGTTCACCCGGGTGTGCAGAAATGGG	4082
QY	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheIleCysAspCysProSerGlyAspPhe	1360
Db	4083	GGCACTGTGCAACTCTGCTGGTGGCGGTTTCAAGTGCAGTTGCCACTTGGAGACTTC	4142
QY	1361	GlyIleProTyrCysGluValThrThrArgSerPheProIleHisSerPheIleThrPhe	1380
Db	4143	GAGAAAGCCCTACTGCGCAGGTACACGCGCAGCTTCCCGCCCACTCTTCATCACTTT	4202
QY	1381	ArgGlyLeuArgGluArgPheHisPheThrLeuIleLeuSerPheIleThrIleGluArg	1400
Db	4203	CGCGGCTCGCGCAGCGTTTCCACTTCACTCGCGCTCTGTTGGCCAGAAAGAGGCG	4262
QY	1401	AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIleHisAspPheValIleLeuGlu	1420
Db	4263	GACGGGTTGCTGTTGTAACAATGGGGGTTCAATAGAAAGCAATGACTTTGTGGCCTGAG	4322
QY	1421	ValIleGlnGluGlnValGlnLeuThrPheSerIleGlyGlySerThrThrValSer	1440
Db	4323	GTGAATCCAGAGCAGGTCCAGTCACTTCTCTGAGGGAGATCAACACCAAGGTGTC	4382
QY	1441	ProPheValProGlyGlyValSerAspGlyGlnTyrHisThrValGlnLeuTyrTyr	1460
Db	4383	CCATTCTGTCGCCGAGAGATCACTGATGGCCAGTGGCATTAAGGTCACTGAATATCAC	4442
QY	1461	AsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnIleValIle	1480
Db	4443	AATAAGCCACTGTTGGGTCAAGCAGGGCTCCCAACAGGCCCATCAAGACAAAGGTGCT	4502
QY	1481	ValValThrValAspGlyCysAspThrGlyValIleAlaLeuArgPheGlySerValLeuGly	1500
Db	4503	GTGGGACCGGTGAATGGCTGTGAACAAGAGTGGCTTGGCGTTGTGAATGTCTCTGGGC	4562
QY	1501	AsnTyrSerCysAlaAlaGlnGlyThrGlnGlySerIleGlySerLeuAspLeuThr	1520
Db	4563	AACTACTCTGTGCTGCCACAGGCGACCCAGGGGTGCACAAAGAACTCTGGATCTGACG	4622
QY	1521	GlyProLeuLeuLeuGlyGlyValProAspLeuProGlyIleSerPheProValArgMetArg	1540
Db	4623	GGGCCCCGTGACTAAGAGGGGGGTGCTTACCTGCCCGAGAGCTTCCAGTCCGAATGGG	4682
QY	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
Db	4683	CAGTTCGTGGGTGCAATCGGAACCTTCAGGTGGACACCCGGCACATAGACATGTGTAC	4742
QY	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIleGlyAsnValCysAspSer	1580
Db	4743	TTCAATGGCAACAAGGACCGGTGCTGCTGCCTCTGCAAGAAAGAAAGTGTGTGACAGC	4802
QY	1581	AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTyrAspAlaPheSerCysGlyCys	1600
Db	4803	AACACTTGCACAAATGGGGGCACTTGGCTGGAACAGTGGACGCTTACGCTCGCAAGTGC	4862
QY	1601	ProLeuGlyPheGlyGlyIleYleSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu	1620

Db	4863	CCCTGGGCTTTGGGGGCAAGACTGGGCCCAAGAAATGGCCAAATCCAGACACTTCTG	4922
Qy	1621	GLYserSerLeuValAlaTrpHisGLYLeuSerLeuProIleSerGlnProTrpTyrLeu	1640
Db	4923	GGCAGCAGACCTGGTGGGCTTGCAATGGCTCTCCGCTGCCCAATCTCCAAACCTGGTACTTC	4982
Qy	1641	SerLeuMetPheArgThrArgGlnAlaArgGlyValLeuLeuGlnAlaIleThrArgGly	1660
Db	4983	AGCCTCATGTCTCCCAAGCGCCAGAGCGGACGAGTCTCGCTGCAGAGGCATCACAGGGGG	5042
Qy	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyYHISValMetLeuSerValGlnGlyThr	1680
Db	5043	CGAGACACCAATCACTTACAGCTACAGAGGGGCCAGCGATGCTGAGCGTGAAGGCACAA	5102
Qy	1681	GlyLeuGlnAlaSerSerLeuValTyrLeuGlnLysProGlyTrpAlaAsnArgLysArgTrpHis	1700
Db	5103	GGGCTTACGGCTCTCTCTCTCGCTGGAGCAGAGCGGGGCCTAAATGACGGTACTGGGAC	5162
Qy	1701	HisAlaGlnLeuAlaLeuGlnValAsnSerGlyLysProGlyHisAlaIleLeuSerPheArg	1720
Db	5163	CATGCACAGCTGGGACCTGGAGCCAGCGGGGGGCTGGCATGCACTTCTGCTTGGAT	5222
Qy	1721	TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGLYLeuHisLeuSer	1740
Db	5223	TATGGGCGCAGAGACGAGGGGGCAACTGGGGCCCCGGCTGCATGTCTGCACCTAGC	5282
Qy	1741	AsnIleThrValGlnGlyLysProGlyProAlaGlyGlyValAlaAsnGlyPheArgGly	1760
Db	5283	AACATACAGTGGGGGGAAATCTTGGGCGACCGCGGGTGTGGCCCTTGCTTGGGGGC	5342
Qy	1761	CysLeuGlnGlyValArgValSerArgThrProGlnGlyValaAsnSerLeuAspProSer	1780
Db	5343	TGTTTGCAGGGGTGTGGGGGTAGCGGATACGCCAAGGGGGCTTAAACACTGGATCCACGC	5402
Qy	1781	HisGLYGLYSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
Db	5403	CATGGGAGACATACACGTGGAGCAAGGCGTGATAGCTGACCTTGAGACTCAAC	5462
Qy	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysArg	1820
Db	5463	CCGTGCTCGCTTAACAGCTAATTCAGCAACGACTGGGACAGCTAATTCCTGACGCTGGAT	5522
Qy	1821	ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisGln	1840
Db	5523	CCAGGTATCTATGGTGTACAACTGTACAAATGTGTGTACCTGAACCCGGTGTACACACAG	5582
Qy	1841	SerValCysThrArgLysProSerAlaProHisGLYTrpThrCysGlnCysProProAsn	1860
Db	5583	TCTGTGTATACCCCGAACGCCCAAGTCCCCCATGGCTATACCTGCAATGTCCCCCAAT	5642
Qy	1861	TyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysProArgGlyTyrTrpGly	1880
Db	5643	TACCTTGGGCGCATCTGTGAGACACAGATATGACACAGCTTGTCCCGGTGGTGGGGGA	5702
Qy	1881	HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProArgCysAsn	1900
Db	5703	CATCCACATGTGGCCCAATGCACCTGTGATGTCCAGAAAGCCTTTGACCAAGACTGCAC	5762
Qy	1901	LysThrSerGlyGlnCysHisCysValGlnLysAsnHisArgTrpProGlySerProThr	1920
Db	5763	AAGACAAAGCGCGAGTGCACCTGCAGAGGAAACACTACCGGCCCCCAGGACGCCCAAC	5822
Qy	1921	CysLeuLeuCysAspArgLysTrpProThrGlySerLeuSerArgValCysAspProGlnAsp	1940
Db	5823	TGCTCTTGTGTACTGTCTCCCAAGAGCTCTTGTCCAGAGTGTGACCTTGAAGAT	5882
Qy	1941	GlyGlnCysProCysLysProGlyValIleGlyValArgGlnCysAspArgCysAspAsnPro	1960
Db	5883	GGCCAGTGTCCATGCACAGCCAGGTGTATCATGGGGGTCACTGTGACCGCTGTGACAACTT	5942
Qy	1961	PheAlaGlnValIleThrThrAsnGlyCysGlnValAsnTrpAspSerCysProArgAlaIle	1980
Db	5943	TTTGTGTAGGTACCAACCAATGGCTGTGAGATGTAAATTATGACACTGTGCCCAAGAGCATT	6002

[illegible]

[illegible]

Alignment Scores:	0
Pred. No.:	2837.00
Score:	100.00%
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	97.06%

Length:	8871
Matches:	2837
Conservative:	0
Mismatches:	0
Indels:	0

DB:	9	Gaps:	0
US-09-916-849A-3 (1-2923) x US-09-788-711A-1 (1-8871)			
QY	1	MechArgSerProAlaThrGlyValProLeuProThProProProProLeuLeuLeu	20
Db	1	ATGGAGAGCCCGGCGCAACGGCGTCCCTCCCAACGCGCGCGCGCTGCTGCTG	60
QY	21	LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu	40
Db	61	TTCGTGCTGCTGCTGCGCGCGCACTTATGGAGAACCAATGGGGCCCTGCTGCTTG	120
QY	41	GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTrpLeuCysProSer	60
Db	121	GGGTCCAGGGAGCGAGGCTCTTGGGGGGCGCTGGCCCCCATGGGCTGCTGCTCATCC	180
QY	61	SerAlaSerAsnLeuTrpLeuTyThrSerArgCysArgAspAlaGlyTyThrGluLeuThr	80
Db	181	TCAGCCGTGCAACCTCTGCTGCTTCAACAGCCGCTGCGAGGATGGCGGCACTGAGTCACT	240
QY	81	GlyHisLeuValProHisHisAspGlyLeuArgValTrpCysProGluSerGluAlaHis	100
Db	241	GGCCACCTGGTACCCCAACAGATGGCTTGAAGGTTTGGTGTCCAGATCCAGAGCCCAT	300
QY	101	IlleProLeuProProAlaProGluGlyCysProTrpSerCysArgLeuLeuGlyIleGly	120
Db	301	ATTCCTCTTACCAACAGCTCTGAAAGGCTGCCCCCTGGAGCTGTGGCTCTTGCGCATTTGA	360
QY	121	GlyHisLeuSerProGluGlyValSerLeuThrLeuProGluGlnHisProCysLeuVala	140
Db	361	GGCACCTTTCCCAACGGGAGAACTCACTGCGCGAGAGACCCGTCCTTAAAGGCT	420
QY	141	ProArgLeuArgCysGlnSerCysIleLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu	160
Db	421	CCACGGCTCAGATGCCAGTCTCTGGAAGCTGGCAAGGCCCCCGGCTCAAGGGAGGGAA	480
QY	161	ArgSerProGluGluSerLeuGlyValArgArgValArgAsnValAsnThrAlaProGln	180
Db	481	AGGTCAACAAAGAGTCCCTGGGGGGCGTCGGAAAAGAAATGTAATATACAGCCCCCAG	540
QY	181	PheGlnProProSerTyArgAlaAlaThrValProGluAsnGlnProAlaGlyThrProVal	200
Db	541	TTTCAAGCCCCCAGACTCAAGGCCACAGTCCCGAGAACCAAGCAGCAGGACCCCTGTT	600
QY	201	AlaSerLeuArgAlaIleAspProArgProGluGlyAlaGlyArgLeuGlyTyThrMet	220
Db	601	GCAATCCCTGAGGGCCATCGAACCCGGAGAGGGTGAAGCAGTCACTGAGATCACCATG	660
QY	221	AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla	240
Db	661	GATCCCTCTTGAATACCGCTCCACCAAGTTCTTCTCCGAGCCCAATCACTGGTGA	720
QY	241	ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr	260
Db	721	GTAACCAACAGCGAGAGCTGGATCGTAGAACCAAGACACCACGCTTCAGGGGTCAAG	780
QY	261	AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuTrpIleLeuValThr	280
Db	781	GCGAGAGACCAAGGCAATGCCCGCAAGAGTGCCTGTGCTACATCCACATCTTGTTACT	840
QY	281	AspThrAsnAspHisAspProValPheGluGlnGlnGlnLysTyArgGluSerLeuArgGlu	300
Db	841	GACACCAATGACATGACCTCTGTGTTGAGAGCAGGAGATCAAGAGAGCTCAGGGAG	900
QY	301	AsnLeuGluValGlyTyArgValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro	320
Db	901	AACCTGGAGGTGGCTATGAGGTGCTCACTGTCAAGGCGCACGAAATGAGAGCCCTCC	960
QY	321	AsnAlaAsnIleLeuTyArgLeuLeuGluGlySerGlyGlySerProSerGluValPhe	340
Db	961	AATGCGAATATTCGTATCCGCTGCTGAGAGGGGTCTGGGGGCAAGCCCTCGAAGTCTTT	1020
QY	341	GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal	360

```
Db 1021 GAGATGACCTCTGCTGGGGATCCGAAACCGGCGCTGTGATCGGGAAGAGTG 1080
Qy 361 GJuseTyrGlnuThrValGluIleAspArgGlnGlyArgAspProGlyProAspSer 380
Db 1081 GAATCTTCAAGCTGACGGTAGAGGCAAGAGACCAAGGTCGGGACCGGGGCTCTCGAAG 1140
Qy 381 ThrThrAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPheSer 400
Db 1141 ACCACGCGCGCTGTTTCTTCTTCTGTGGAGATGACAAATGATATGCCCCCGAGTTAAT 1200
Qy 401 GJuvAspTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
Db 1201 GGAAGCGGTATGTGTCCAGGTGAGGAGATGATCTCCAGGGGCCCAAGTACTCCAA 1260
Qy 421 ValThrAlaSerAspArgAspGlySerAspAlaValAlaIleTyrSerIleMetSer 440
Db 1261 GTCAAGCTCGGATCGAGACAGAGGAGCAATGCGGTGTGCACTATAGCATCATGAGT 1320
Qy 441 GJyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
Db 1321 GCGAATGCTCGGGGACAGTTTATCTGATGCCCAAGCTGAGCTCTGATGTGGTGAAC 1380
Qy 461 ProLeuAspTyrGlnThrThrLysGluTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db 1381 CCTCTTGACTATGAGACGACCAAGAGATACACCTTACGGGTGCGACACAGATGTGTGC 1440
Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db 1441 CATTCCCACTCTCTATATGCTCTGGCTGTGTGACAGTACAGGTCTGTGATTCACAGAC 1500
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db 1501 AATGCCCATCTCTGTGTGACAGACCCCTTCCAGGCTACTGTCTGTGAGAGCGTCCCTTA 1560
Qy 521 GJYTYrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db 1561 GGCCTACCTGGTCTCATGTCCAGGCTATCGACGCTGATGCGTGTGACATGCGCCCTG 1620
Qy 541 GJUTYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db 1621 GAATACCGCTTGCTGGGGTGGGACATGACTTCCCTTCAACATCAACATGGCACAGGC 1680
Qy 561 TTPDLeuSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db 1681 TGGATCTCTGTGGCTGTAACCTGAGCGGAGCGAAGTGAATTCTTCAAGCTTTGGGGTA 1740
Qy 581 GJuaIlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThrVal 600
Db 1741 GAAGCTCGAGACCATGCACTCCAGCATCACTGCTCGGCGCAAGTGCAGCTGACTGTC 1800
Qy 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db 1801 CTGGATGTCAAGACCAACATCAACCTTTACCAACCAAGATCAACATGGCGGCTCAAT 1860
Qy 621 GJuaAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db 1861 GAGATGACGCTGTGGGACCAAGCGTGTGACGGGTGCTGAGTGTGAGCGTATGTCTAT 1920
Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db 1921 AATGTCATCACTTACCAATCAACAGTGCATATCTCGAAGACCGCTTCTCCATCAACAGC 1980
Qy 661 GJnSerGlyGlyValLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGln 680
Db 1981 CAAAGTGGTGGGTGTATCTCTTGTGCTGCTGCACTGCACTCAAACTTAAAGGGGCA 2040
Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db 2041 TATGTGTGGCTGTATACCGCTCCGATGCGCATCGGCAAGACACGGCACAATGTGGTG 2100
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 2101 AATGTACCGACGCAACACCAACATGCTCTTCTTTCAGAGCTCCCACTATACATGAT 2160
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValIleIleSerAlaThrAspGluAsp 740
Db 2161 GTTAATGAGACCGGCGGCAAGCAACAGGTGGTGTATGACGGCCACGATGAGGAC 2220
Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACGGTGAGAAATGCCCGATCACTTCACTTCAATGAGAGACAGATCCCCAGTTCCGCATC 2280
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db 2281 GATGCAGACACGGGGCTGTCCACCAAGGCTGAGTGCATACAGAGACCAAGTGTCT 2340
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db 2341 TACACCTTGGCCATTAATGCTCTGGGACAAATGCAATTCCTCCAAAGTCCGACACCTAC 2400
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db 2401 CTGAGATCTGTGTGAACGATGATGACAAATGCAATGCCCTCAGTTCTGTGAGATCTTAC 2460
Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGCGAGTGTCTATGAGAGATGTGCCACCTTCACTACGCTCTGCAATCTCACCTCT 2520
Qy 841 AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyValAspAspGly 860
Db 2521 GATGTGATTTCTGACCTTAAATGACAGGCTTCTTACACCTTCCAAAGAGGCAACATGGA 2580
Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GACGCTGACTTATATGTGATGCCACGTCAAGGCATCGTGCAAGCTACGAGGCTGGAT 2640
Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyGlyMetProPro 900
Db 2641 CGAGGAAACGTGGCCCAATATGTCTTGGCGGATGTGACAGTGAACAGAGGATGTCCCCA 2700
Qy 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db 2701 GCCGCAACCTTATGGAAGTACATCACTGTGTGATGTGAATGACAAATCCCTGTGC 2760
Qy 921 PheGluIleAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTGAGCAGATGAGTTGATGTGTGTGGAAGAGAACAGGCCATTTGGGCTACCGCTG 2820
Qy 941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db 2821 GCCCGGTCAACAGCCATGACCCCGATGAAAGCACCAATGCCCAGATTATGTACCAAGAT 2880
Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 2881 GTGAGGGCAACATCCCTGAGGCTTCCAGCTGACATCTTCCGGGAGCTGACAGCC 2940
Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGGTACACTTACCTAGAGGACCGGCGCTGATGACGTCCTGTGATCAAGGCGACGCA 3000
Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspAsnPro 1020
Db 3001 GCTCCTGTGTGAGCGGGGCTTACAGTCCAGTCCGCTTGTGACCGGAATGACAAACCA 3060
Qy 1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAspArgSerSerSer 1040
Db 3061 CCAAGTGTGGGCAACTTGTGATCTTTTCAACAACATATGTACCAATCGCTCAAGACGC 3120
Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3121 TTCCCTGGGGGTGCATTGGCCGAGTACTGCTCCATGACCTGTATCTCAGATGTGTG 3180
Qy 1061 ThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuAsnAlaSerThrGly 1080
Db 3181 ACTTACAGCTTGAAGCGGGGAAATGAATCAAGCTGTGTCTCAATGCTTCCACGGGT 3240
```

QY 1081 GIULeuLYLeuSerArgAlaLeuAspAsnAsnArgProLeuGluValIleMetSerVal 1100  
DB 3241 GAGCTGAAGCTTAAGCGCGGCACTGACCAACAACCGGCTCTTGAGGCGCATCATGAGCGTG 3300  
QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
DB 3301 CTGGTGTCAAGCGGCTTACACAGCTGACCGGCCAGTGGCGGCTGTGTGACCATCATC 3360  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
DB 3361 ACCGATGAAGTGTCTTCCACACACATCACCTGCGCTGGAGGACATGTCAACCGAGCGG 3420  
QY 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
DB 3421 TTCCTGTCAACACTGAGGCTCTTTCATTCAGCGGTTGGCGCGCAAGCTGGCAAGCCA 3480  
QY 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyValIleIle 1180  
DB 3481 CCGGACCAAGTGTGTCTTCAACAGTCAAGCGGAGACACGACCGCCCGGCGGCCCATC 3540  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyValGlyProProPheLeu 1200  
DB 3541 CTCAACTGAAGCTGTGCGGTGGGCGAGCGGCCAGGCGCCGGGGCGGCGCTTCTCTG 3600  
QY 1201 ProSerGluAspLeuGlnGluArgLeuTYrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
DB 3601 CCTCTGAGGACCTGACAGAGCGCTATACCTCAACCGGACCTGTGACAGGCGCATCTCG 3660  
QY 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTYr 1240  
DB 3661 GGCACAGCGGTGTGCTTCCCTTCAAGACCAATCTGTGCGGAGGCCCTGCGGAACTAC 3720  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
DB 3721 ATCCGCTGCGTGTGCGTGTGCTGCGCTTCACTCCGCGCTTCAATCGCTCTCTCTC 3780  
QY 1261 ValLeuPheArgProIleHisProValGlyValLeuArgCysArgCysProProGlyPhe 1280  
DB 3781 GTGCTCTTCCGCGCCATCCACCCCGTGGAGGCTCGCTGCGCGCGCGCGCGCTCTC 3840  
QY 1281 ThrGluAspPheTYrCysGluThrGlnValAlaPheLeuCysTYrSerArgProCysGlyProHis 1300  
DB 3841 ACCGAGTACTACTGACAGACCGAGGTGACCTTGTGCTGCTGCGCGCTGTGGCCCCAC 3900  
QY 1301 GlyArgCysArgSerArgGluGlyTYrTYrCysLeuCysArgAspGlyTYrThrGly 1320  
DB 3901 GGGCGCTGCGCGGCGCGGCGGCGGCTACCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 3960  
QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIlyAsnGly 1340  
DB 3961 GAGCAGTGTGAGGTGAGTGTCTCCCTCAGCGCGCTTGGACCCCGGAGTGTGTGCAAGATGGG 4020  
QY 1341 GluThrCysValAlaAsnLeuValGlyGlyPheLeuCysAspArgProSerArgIlyAspPhe 1360  
DB 4021 GGGACCTGTGTCAACCTGCTGTGGGGGCTTCAAGTGTGATGTGCCATCTGAGACTTC 4080  
QY 1361 GluIlyProTYrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
DB 4081 GAAAGAGCCCTACTGCCAGCGGACCAAGCGGAGCTTCCCGGCCCATCTTCAATCACTTT 4140  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIlyGluArg 1400  
DB 4141 CGGGGCTGGCGCAGGGTTCACCTTCACTTCCCTCTCGTTTGCACAAAGGAGCGG 4200  
QY 1401 AspGlyLeuLeuLeuTYrArgGlyArgPheAsnGlyIlyHisIlyAspPheValAlaLeuGlu 1420  
DB 4201 GAGCGGT 4260  
QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGlyLeuSerThrThrValSer 1440  
DB 4261 GTGATCCAGAGAGGCTCAGCTCACTTCTGTGAGGGAGTCAACCAACGAGTGTCC 4320

QY 1441 ProPheValProGlyValValSerAspGlyGlnThrHisThrValGlnLeuLeuTYrTYr 1460  
DB 4321 CCATTCGTGGCCGAGAGAGTCAAGTGAATGGCCAGTGGCATTAAGGTGCAGCTGAATATCTAC 4380  
QY 1461 AsnIlyProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIlyValAla 1480  
DB 4381 AATAAGCCACTGT 4440  
QY 1481 ValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
DB 4441 GTGTGACCGT 4500  
QY 1501 AsnTYrSerCysAlaAlaGlnGlyThrGlnGlySerIlyIlySerIlySerLeuAspLeuThr 1520  
DB 4501 AACTACTCTGT 4560  
QY 1521 GlyProLeuLeuLeuGlyValProAspLeuProGlyIlySerPheProValArgMetArg 1540  
DB 4561 GGGCGCTGT 4620  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
DB 4621 CAGTTCGTGGCTGT 4680  
QY 1561 PheIleAlaAsnAsnGlyTYrValProGlyCysProAlaIlyIlyAsnValCysAspSer 1580  
DB 4681 TTCATTGTCCAACAATGGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4740  
QY 1581 AsnThrCysHisAsnGlyTYrThrCysValAsnGlnThrPheAspAlaPheSerCysGlyCys 1600  
DB 4741 AACACTGTGCACAAATGGAGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
QY 1601 ProLeuGlyPheGlyGlyIlySerCysAlaGlnIlyMetAlaAsnProGlnHisIlyPheLeu 1620  
DB 4801 CCCCTGGGCTTTGGGGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4860  
QY 1621 GlySerSerLeuValAlaIleThrHisGlyLeuSerLeuProIleSerGlnProTYrIleu 1640  
DB 4861 GGCACAGCGT 4920  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
DB 4921 AOCCTCATGTTCCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4980  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
DB 4981 CGCAGACACATCAACCTTACAGTACAGAGGGGCGAGTGTGTGTGTGTGTGTGTGTGTGT 5040  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAlaAsnAspGlyAspThrHis 1700  
DB 5041 GGGCTTCAAGGCT 5100  
QY 1701 HisAlaGlnLeuAlaLeuGlnValAlaSerGlyProGlyHisAlaIleLeuSerPheAsp 1720  
DB 5101 CATGCAAGCTGT 5160  
QY 1721 TYrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyIlyLeuHisLeuSer 1740  
DB 5161 TATGGGCAAGAGAGAGAGAGGCAACTGTGGGCCCGCGGCTGTGTGTGTGTGTGTGTGT 5220  
QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly 1760  
DB 5221 AACATTAACGT 5280  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAlaSerSerLeuAspProSer 1780  
DB 5281 TGT 5340  
QY 1781 HisGlyIlySerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
DB 5341 CATGGGAGAGCATCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400  
QY 1801 ProCysProAlaAsnSerTYrCysSerAsnAspTrpAspSerTYrSerCysSerCysAsp 1820



Db	5401	CCGTGCTCTGCTAAGCTATTGACAGAAAGCTGGAGACAGCTATTCTGACGCTGTGAT	5460	Db	6481	ACCATGTCAGCGCCCAACATTGTCACTCCGTAGTGCCTTGGAACAAAGGAACCTTGCT	6540
Qy	1821	ProglYTYTTrgLYAspAsnCYsThrAsnValCYAspLeuAsnProCYsGluHisGln	1840	Qy	2181	GlyAlaLeuLeuProArgTYTTrgLYuAlaLeuArgGlyGluGlnProProAspLeuGluThr	2200
Db	5461	CCAGGTACTATGAGTGAACAAGTACTATATGTGTGACTGAACCCGTGTAGACACAG	5520	Db	6541	GGGGCCAAAGCTGCCCGCTTAGAGGCTCGCTGGGGAGAGAGCCCCCGGACTTAGACA	6600
Qy	1841	SerValCYsThrArgLYsPProSerAlaProHisGlyTYTTrgLYuCYsGluCYsProProAsn	1860	Qy	2201	ThrValILEuProGluSerValPheArgGluThrProProValValArgProAlaGly	2220
Db	5521	TCGTGTGTATCCCGCAAGCCCAAGTGCCTCCCATGTGCTATACCTGACGAGTGTCCCAAT	5580	Db	6601	ACAGTCATTGCTGAGTGTCTTTCAGAGAGAGCCCCCGTGTAGGCCCGCAGGC	6660
Qy	1861	TYrLeuGLYProTYrCYsGluThrArgGluLeuAspGlnProCYsProArgGlyTYTTrgLY	1880	Qy	2221	ProGlyGluAlaGlnGluProGlyGluLeuAlaArgGlnArgHisPProGluLeu	2240
Db	5581	TACCTTGGGCTACACTGTAGACACAGGATTGACACAGCTTGTCCCGTGGCTGGGGGA	5640	Db	6661	CCCGAGAGGCCCCAGAGACAGAGAGCTGGCAACGGCCACAGCCAGCCCGGAGCTG	6720
Qy	1881	HisProThrCYsGlyPProCYsAsnCYsAspValSerLYsGlyPheAspProAspCYsAsn	1900	Qy	2241	SerGlnGlyGluAlaValAlaSerValILETYrArgTYrAsnValAGLYLeuLeuPro	2260
Db	5641	CATCCCAAGTGTGGCCCAATGCAACTGTATGTACAGAAAGGCTTTGACCAACCTGCAAC	5700	Db	6721	AGCCAGGCTGAGGCTGTGGCCAGGCTTCATCTACCCGACCTTGCCGGGCTACTGCTT	6780
Qy	1901	LYsThrSerGlyGlyCYsHisCYsLYsGluAsnHisTYrArgProProGlySerProThr	1920	Qy	2261	HisAsnTYrAspProAspLYsArgSerLeuArgValProLYsArgProILELeuAsnThr	2280
Db	5701	AAGACAAAGCGGAGTGCACCTGCAAGAGAAACCACTACGGGCCCCAGGAGCCCAACC	5760	Db	6781	CATAACTATGACCCCTGACAGCGCAGCTTGAGAGTCCCAAAACGCCGATCATCAACA	6840
Qy	1921	CYsLeuLeuCYsAspCYsTYrProThrArgLYsSerLeuSerArgValCYsAspProGluAsp	1940	Qy	2281	ProValValSerILESerValHisAspAspGlyGluLeuLeuProArgAlaLeuAspLYs	2300
Db	5761	TGCTCTGTGTGTGACTGTACCCCAAGGCTCTCTGTCCAGAGTCTGTGACCTTAGAGAT	5820	Db	6841	CCCGTGTGACATCAAGGCTCATGATGATGAGAGCTTGTGCCCGGGCCCTGGACAA	6900
Qy	1941	GlyGlnCYsPProCYsLYsPProGlyValILEGlyArgGlnCYsAspArgCYsAspAsnPro	1960	Qy	2301	ProValThrValGlnPheArgLeuLeuGluThrGluGluArgTYrLYsProILECYsVal	2320
Db	5821	GGCCAGTGTCCATGACAGCCAGGTGTCAACGGGCGTCAAGTGTGACCGCTGTACAACTC	5880	Db	6901	CCCGTCAAGGTGACGTTCCGCTGTGGAGACAGAGAGCCGAGCCAGCCATCTGTCTC	6960
Qy	1961	PheAlaGluValThrThrAsnGlyCYsGluValAsnTYrAspSerCYsProArgAlaILE	1980	Qy	2321	PheTYrAsnHisSerILELeuValSerGlyTYrGlyTYrPProSerAlaArgGlyCYsGlu	2340
Db	5881	TTTGTGTGAGTCAACCAATGCTGTGTGAAGTGAATTATGACAGCTGCCCCAGACGAAAT	5940	Db	6961	TTCTGGAACCAATTCAATCTGTGTGAGTGCACAGGTGGCTGTGGCCAGAGGCTGTGA	7020
Qy	1981	GluAlaGlyILETYrProArgThrArgPheGlyLeuProAlaAlaProCYsPro	2000	Qy	2341	ValValPheArgAsnGluSerHisValSerCYsGlnCYsAsnHisSerThrSerPheAla	2360
Db	5941	CAGGCTTGGGATGTGGTGGCCCCGATACCCGCTTGGGCTGCTGCTGCTCCCTGTCCC	6000	Db	7021	GTCGCTTCCGCAATGAGAGCCAGTCAAGTGCAGAGTCAACCAATGACAGCTTGCT	7080
Qy	2001	LYsGlySerPheGlyTYrAlaValArgHisCYsAspGluHisArgGlyTYrLeuProPro	2020	Qy	2361	ValLeuMetAspValSerArgArgGlyuAsnGlyGluILEuProLeuLYsThrLeuThr	2380
Db	6001	AAAGGCTCTTTGGGACGTGCTGGGCCACTGTGATGAGCAAGAGGGGTGGCTCCGCCA	6060	Db	7081	GTCGTCATGAGACGTTTCTCCGGCGGAGAGATGGGAGATCTGCCACTGAAACACTGAC	7140
Qy	2021	AsnLeuPheAsnCYsThrSerILEThrPheSerGluLeuLYsGlyPheAlaGluArgLeu	2040	Qy	2381	TYrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu	2400
Db	6061	AACCTCTTCAACTGACGTCACATCACTCTCTCAGAACTGAAAGGCTTGCCTGACGGCTA	6120	Db	7141	TACGTGGCTTAGGGTGCACCTTGCTGCCCTCTGTGCTCACTTCTCTCTCACTCTC	7200
Qy	2041	GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg	2060	Qy	2401	LeuArgGluLeuArgSerAsnGlnHisGlyILEArgArgAsnLeuThrAlaLeuGly	2420
Db	6121	CAGCGGAATGAGTCAAGGCTTAGACTCAAGGCGCTCCCAAGCAAGCTTCTCTGCGAC	6180	Db	7201	TTGGGATCTGTGCCTCCCAACCAACAGGCAATCCGAGCTAACCTGACAGCTGCTGGGC	7260
Qy	2061	AsnAlaThrGlnHisThrAlaGlyTYrPheGlySerAspValLYsValAlaArgGlnLeu	2080	Qy	2421	LeuAlaGlnLeuValPheLeuLeuGlyILEAsnGlnAlaAspLeuProPheAlaCYsThr	2440
Db	6181	AACGCCACGACAGACACAGCTGGTACTTCGGCAGAGACGTCAAGSTGGCTTACACAGCTG	6240	Db	7261	CTGGCTCAGCTGGTCTTCTCTCTGGGAATCAACAGGCTGACCTCCCTTTGGCTGACA	7320
Qy	2081	AlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln	2100	Qy	2441	ValILEalILEuLeuLeuHisPheLeuTYrLeuCYsThrPheSerTYrAlaLeuLeuGlu	2460
Db	6241	GCCAGCGCGCTGTGCCCCACAGAGACCCACGCGGGCTTGTGGCTGTCTGCCACACAG	6300	Db	7321	GTCATTGCAATCCGTGCTGACCTTCTGTACCTCTGCACTTTTCTCGGGCTGTCTGGAG	7380
Qy	2101	AspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120	Qy	2461	AlaLeuHisLeuTYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Db	6301	GACGTGCACTTCACTGAGAAATCTGCTGGGAGTGGGAGCGCCCTCTGTGACACACACAC	6360	Db	7381	GCTTTCACCTGTACCGGGCACTCACTGAGGTGGCAATGTCAACACCGGCCCCATGGGC	7440
Qy	2121	LYsArgHisTYrPGLYLeuILEGlnGlnThrGluGlyTYrAlaATrPLeuLeuGlnHis	2140	Qy	2481	PheTYrTYrMetLeuGlyTYrPGLYValProAlaPheILEThrGlyLeuAlaValAGLYLeu	2500
Db	6361	AAGCGGCACCTGGGACGTGATCCAGACAGACAGAGGTGGCACCGGCTGTGCTTCAGCAC	6420	Db	7441	TTCTACTACATGCTGGGCTGGGCGGTGCTGCTTCATCAACAGGCTTAGCGGTGGCGCTG	7500
Qy	2141	TYrGluAlaTYrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTYrLeuSerProPhe	2160	Qy	2501	AspProGluGlyTYrGlyAsnProAspPheCYsTYrLeuSerILETYrAspThrLeuILE	2520
Db	6421	TATGAGGCTTACGCGAGTGCCTGTGGCCAGAACATGCGGCACTTACTTAAGCCCTTTC	6480	Db	7501	GACCCCAAGGCTTACGAGAACCTTGACTTGTGCGGTCTTCCCACTTATGACACCTCATC	7560
Qy	2161	ThrILEValThrProAsnILEValILESerValValArgLeuAspLYsGlyAsnPheAla	2180	Qy	2521	TYrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTYrILEu	2540
				Db	7561	TGGAGTTTGTGGCCCGGTGGCTTTGGCTTGGCTCGATGAGTGTCTTCTGTACATCTGTG	7620



QY 2541 AAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnuYsYsGlyProValSer 2560  
Db 7621 GCGGCCCGGGCTCTCTGTGCTGCGCAGCGGCAAGGCTTTAGAAAGAAAGTCTGCTCG 7680  
QY 2561 GYLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeu 2580  
Db 7681 GGGCTGAGCCCTTCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7740  
QY 2581 LeuSerValaSerAspThrLeuPheHisTyrLeuPheAlaThrCysAlaCysIle 2600  
Db 7741 CTCTCTCTCAACAGCAGCAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerIysGlnuValArgValaLeu 2620  
Db 7801 CAGGGCCCTTCATCT 7860  
QY 2621 LysLeuAlaCysSerArgIysProSerProAspProAlaLeuThrThrIysSerThrLeu 2640  
Db 7861 AAGCTTGCTGCGAGCCGCAAGCCGAGCCCTGACCTGCTGACCAAGTCCACCTG 7920  
QY 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db 7921 ACCTCTCTCTCAACATCT 7980  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyIysSerGlnProSerTyrIle 2680  
Db 7981 GACTCGCCCGCTCTCTCTGACACACACAGTCTGCGGCAAGTCCAGCCCACTACATC 8040  
QY 2681 ProPheLeuLeuArgGlnuGlnuSerAlaLeuAspProGlyGlnGlyProPogIysLeuGly 2700  
Db 8041 CCTCTCTGCTGAGGAGGAGGCTCGACCTGACCTGACCTGACCTGACCTGACCTGAGG 8100  
QY 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8101 GATCCAGGCGAGCTGTTCTCTGGAAGGTCAAGACCAAGAGATGATCTCTGACAGACTCC 8160  
QY 2721 AspSerAspLeuSerLeuGlnuAspAspGlnuSerGlySerTyrAlaSerThrHisSerSer 2740  
Db 8161 GACAGTGAAGCTGCTTGAAGAGAGACGACGAGTGTCTCTCTCTCTCTCTCTCTCTCT 8220  
QY 2741 AspSerGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnuGlnu 2760  
Db 8221 GACAGTGAAG 8280  
QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGlnuArgLeuProLeuHisSerThrProIysAsp 2780  
Db 8281 TGGGATAGCTGTGCTGGGGCTGTGGAGCAGAGACTGCTCTGACAGTATCTCCCAAGAT 8340  
QY 2781 GlnGlyProGlyProGlyIysAlaProTyrProGlyIysAspPheGlyThrThrAlaYsGln 2800  
Db 8341 GGGGGCCCAAGGCTGTGCAAGGCCCCCTGGCCAGAGACTTTGGGACCAAGCAAAAGAG 8400  
QY 2801 SerSerGlyAsnGlnAlaProGlnuGlnuArgLeuArgGlnuAsnGlyAspAlaLeuSerArg 2820  
Db 8401 AGTAGTGGCAACGGGCGCTCTGAGAGAGGCTGTGGAGATGAGATGAGATGAGATGAGAT 8460  
QY 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisYsGly 2837  
Db 8461 GAGGGGTCCTTGAAGCCCTTCCAGGCTCTTCTGCCAGCTCAAGTCAAGT 8511

RESULT 4  
US-10-292-798-931  
Sequence 931, Application US/10292798  
Publication No. US2003023583A1  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ASURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798

CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 931  
LENGTH: 24370  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
LOCATION: source  
FEATURE:  
LOCATION: (1) .. (24370)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201) .. (3510)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (8553) .. (9200)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11163) .. (11385)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11634) .. (11749)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11929) .. (12017)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (12408) .. (12565)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (12927) .. (13090)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (13263) .. (13388)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (13732) .. (13898)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (14199) .. (14486)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (14575) .. (14749)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (14989) .. (15130)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (15312) .. (15432)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (15856) .. (16039)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (16226) .. (16327)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (17669) .. (17763)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (17972) .. (18181)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18702) .. (18885)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (19001) .. (19129)  
FEATURE:

```

NAME/KEY: CDS
LOCATION: (19231) .. (19411)
FEATURE:
NAME/KEY: CDS
LOCATION: (19545) .. (19712)
FEATURE:
NAME/KEY: CDS
LOCATION: (19815) .. (19941)
FEATURE:
NAME/KEY: CDS
LOCATION: (20054) .. (20254)
FEATURE:
NAME/KEY: CDS
LOCATION: (20547) .. (20721)
FEATURE:
NAME/KEY: CDS
LOCATION: (21048) .. (21160)
FEATURE:
NAME/KEY: CDS
LOCATION: (21338) .. (21417)
FEATURE:
NAME/KEY: CDS
LOCATION: (21507) .. (21633)
FEATURE:
NAME/KEY: CDS
LOCATION: (21721) .. (21843)
FEATURE:
NAME/KEY: CDS
LOCATION: (22399) .. (22526)
FEATURE:
NAME/KEY: CDS
LOCATION: (22761) .. (22849)
FEATURE:
NAME/KEY: CDS
LOCATION: (22954) .. (23148)
FEATURE:
NAME/KEY: CDS
LOCATION: (23287) .. (23457)
FEATURE:
NAME/KEY: CDS
LOCATION: (23557) .. (23791)
FEATURE:
NAME/KEY: CDS
LOCATION: (24143) .. (24170)
US-10-292-798-931
Alignment Scores:
Pred. No.: 0
Score: 1103.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.74%
DB: 12
Length: 24370
Matches: 1103
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-09-916-849a-3 (1-2923) x US-10-292-798-931 (1-24370)
QY 1 MetatgserProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
DB 201 ATGGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 260
QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40
DB 261 TTGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320
QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSer 60
DB 321 GGGTCCAGGGGAGGAGGCTCTTCGGGGGCGCTGCGGCCCAATGGGCTGCTGCTCATCC 380
QY 61 SerAlaSerAspLeuThrLeuThrThrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80
DB 381 TCAGGCTCGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValThrCysProGluSerGluHis 100

```

```

DB 441 GGCCACCTGTATACCCACAGATGCGCTGAGGGTTGTGTCCAGATCCGAGGCCCAT 500
QY 11eProLeuProProAlaProGluGlyCysProThrPserCysArgLeuLeuGlyTleGly 120
DB 501 ATTCCTTACCAACAGCTCTGTAAGGCTGCCCCCTGAGCTGCTGCTGCTGCTGCTGCTG 560
QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGluHisLeuProCysLeuLeuVal 140
DB 561 GGCCACCTTCCCAACAGGGAGGCTCACTGCCCCGAGAGACCCGCTGCTTAAGGCT 620
QY 141 ProArgLeuArgCysGlnSerCysAlaLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160
DB 621 CCACGGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
QY 161 ArgSerProGluGluSerLeuGlyGlyValArgArgGlySerAspValAspThrAlaProGln 180
DB 681 AGGTCCAGAGAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
QY 181 PheGlnProProSerArgGlnAlaThrValProGluHisGlnProAlaGlyThrProVal 200
DB 741 TTCCAGCCCCCAGGTACAGGCCAGTGCCTGAGAGACAGCCAGGACACCCCTGTT 800
QY 201 AlaserLeuArgAlaAlaAspProAspGluGlyGlyAlaGlyArgLeuGlyThrMet 220
DB 801 GCATCCCTGAGAGGCGCATGACCCGAGAGGCTGAGGCGGTGCACTGAGCTACACCATG 860
QY 221 AspAlaLeuPheAspSerArgSerAspGlnPhePheSerLeuAspProValThrGlyAla 240
DB 861 GATGCCCTCTTGTATAGCGGCTCCACCAAGTCTCTCTCCAGAACCATGCTGCTGCTCA 920
QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260
DB 921 GTAACCAAGCGGAGAGGCTGATGCTGAGACCAAGAGCACCAAGTCTTCAAGGCTGACG 980
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrTleLeuValThr 280
DB 981 GCGCAGGACCAAGGAGATGCCCCGAGAGTCCCTGCTGCTACCTCACTGCTGCTGCTGCT 1040
QY 281 AspThrAspAspHisAspProValPheGluGlnGlnGluValLysGluSerLeuArgGlu 300
DB 1041 GACACCAATGACCATAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100
QY 301 AsnLeuGluValGlyValGluValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro 320
DB 1101 AACTGAGGTTGGCTATGAGGTGCTCACTGTCAGGGCCAGATGAGTGGCCCTCC 1160
QY 321 AsnAlaAsnTleLeuThrArgLeuLeuGluGlySerGlyLysSerProSerGluValPhe 340
DB 1161 AATGCCAATATTCGTACCGCTCTGCTGAGGGGCTCTGAGGGGACCCCTGTAAGTCTTT 1220
QY 341 GluTleAspProArgSerGlyValTleArgThrArgGlyProValAspArgGluGluVal 360
DB 1221 GAGATCAACCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280
QY 361 GluSerTyGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380
DB 1281 GAATCTTACAGCTGACGCTGAGGAGGATGACCAAGGAGTCCGAGGCTCTCGAGAGT 1340
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400
DB 1341 ACCACAGCGGCTGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1400
QY 401 GluLysArgTyValAlaGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
DB 1401 GAGAGGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1460
QY 421 ValThrAlaSerAspAlaGlyGlySerAsnAlaValValHisTySerTleMetSer 440
DB 1461 GTCAAGGCTGATGATGACAGAGGAGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1520
QY 441 GlyAsnAlaArgGlyGlnPheTyLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460

```

Db 1521 GGCAATGCTCGGGACAGTTTATCTGATGCCAGACCTGAGTGTGAGC 1580  
 Qy 461 ProLeuaspTyrGluThrThrylsGluTyrThrLeuArgValArgIleaspGly 480  
 Db 1581 CCTCTGACGTATAGAGACAGCAAGAGTACACCTTACGGGTGGAGCACAGATGGTGC 1640  
 Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1641 CGTCCCCCACTCTATATGCTCTGGCTGGTGAACAGTACAGGTCTCTGGATATCAACGAC 1700  
 Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 Db 1701 AATGCCCAATCTTCGACAGACCCCTTTCAGGCTACGTCTCGAGAGTGTCCCTTA 1760  
 Qy 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db 1761 GGTCTACTGGTCTTCATGTCCAGGCTATGACGCTGATGCTGTGACAAATGCCCTG 1820  
 Qy 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db 1821 GAATACCGCTCTGCTGGGTGGAGCATGACTCCCTTCAACATCAATGACACAGGC 1880  
 Qy 561 TrpIleSerValAlaIleGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
 Db 1881 TGGATCTCTGTGGCTGTGTAAGTGAACCGGAGAGAGTGAATTTCTACAGCTTTGAGGTA 1940  
 Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1941 GAAAGCTCGAAGCATGGCATCTCCAGCATCACTGCTCGGCTCGGCTGACGCTGACCTGTC 2000  
 Qy 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 Db 2001 CTGGAGTGTCAACACACAAATCCAACTTACCCCAACAGATACACAGTGGCGCTCAAT 2060  
 Qy 621 GluAspAlaIleValGlyThrSerValValThrValSerAlaValAspAspAspAlaHis 640  
 Db 2061 GAGAGTGCACCTGTGGCACAGGAGTGTGACGGTGTGACCTGTGGACCTGTAGTGCAT 2120  
 Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 2121 AGTGTATCATCACTACCAAGATCAACAGTGCATATCTGAAACGGCTTCTCATCACAGC 2180  
 Qy 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln 680  
 Db 2181 CAAGAGGTGTGGGTGTATCCCTTGCCTGTGCATGTGACATCAAACTTGAGCGGACG 2240  
 Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 2241 TATGTGTGCTGTACCGCTCCGATGACATCGGACATCGGACAGCACGATTTGAGTG 2300  
 Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisIleThrValAsn 720  
 Db 2301 AATGTACCGACCGCAACCCATCGTCTGTCTTCAAGAGCTCCCATATACAGTGAAT 2360  
 Qy 721 ValAsnGlnAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db 2361 GTTAAATGAGACCGGCGGACGACCAACGATGTGATGACGCGCACAGATGAGGAC 2420  
 Qy 741 ThrGlyLeuAsnAlaArgIleThrTyrPheMetGlnAspSerIleProGlnPheArgIle 760  
 Db 2421 ACAGGTGAGATGCCCGCATCACTTCTATGAGAGACAGCATCCCGCATTCGGATC 2480  
 Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db 2481 GATGACAGACAGGGGGCTGTCAACCAAGCTGAGCTGATATGAAACCAAGTGTCT 2540  
 Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerSerAspThrThrTyr 800  
 Db 2541 TACACCTGTGGCATTAATCTCGGAGCAAAATGGCATTCCTCCAGAGTCCGACCACTTAC 2600  
 Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyr 820  
 Db 2601 CTGGAAGATCTCTGGTGAACGAGTGAATGACAAATGCCCTCAAGTTCTCTGGAGACTCTTAC 2660

Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2661 CAGGCGACGTGTATGAGAGATGTGCCACCTTCACTACAGTCTCTGACATCTCAACGACT 2720  
 Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2721 GATCGTGAATTCGAGCTTAATGTGCGAGGTCCTTCAACCTTCCAGAGGCGCACATGGA 2780  
 Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 Db 2781 GACGTGACTTATATGTGTGATCCAGTCAAGCATGTGCGAACCTTACGAGGCTGGAT 2840  
 Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro 900  
 Db 2841 CGAGAGAACGTGGCCAGATATGTCTTGGGGGACATGACGTGACAGAGGATGCCCCCA 2900  
 Qy 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920  
 Db 2901 GCCCGCACACTATGGAAGTGAACGTCACTGTGTGGAATGTGAATGACAAATCCCCCTGTC 2960  
 Qy 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 Db 2961 TTGAGCAGATGATGTGTGATGTGTGTGGAAGAGAACGCCCATTTGGGCTTACGCTG 3020  
 Qy 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 3021 GCCCGGTCAAGCACCATGACCCCGATGAGGACCAATATGCCAATATATGATACAGATT 3080  
 Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla 980  
 Db 3081 GTGGAGGGGCAATATCTGAGAGTCTTTCAGCTGGAATCTTCTCGGGAGCTGACAGCC 3140  
 Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000  
 Db 3141 CTGTGAGACTTATGACTTACAGAGACCGGCTGAGTACGTCTGTATCCAGGCTCAAGTGA 3200  
 Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 3201 GCTCCTGTGTGAGCGGGGTCAAGTCCAGTCCCGCTCTTGAACCGCAATCAACACCA 3260  
 Qy 1021 ProValIleLeuGlyAsnAspGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040  
 Db 3261 CCAAGTCTGGGCAACTTGAATCTTTCACCAACTATGTACCAATGCTCAAGCGAC 3320  
 Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 3321 TTCCCTGGGGGTGCCATTTGGCCGAGTACTGCTCCATGACCTGATATCTCAATATGTCG 3380  
 Qy 1061 ThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3381 ACTTACAGCTTGTGACCGGGGAATGAACTCAACCTGTGCTCTCAATGCTTCCACGGGT 3440  
 Qy 1081 GluLeuValSerSerArgAlaIleAspAsnAsnAspProLeuGlnAlaIleMetSerVal 1100  
 Db 3441 GAGCTGAAGCTTAAGCGCGCATGTGACAAACCGGCTCTGAGGCGCATATGAGCGTG 3500  
 Qy 1101 LeuValSer 1103  
 Db 3501 CTGGTGTCA 3509

RESULT 5  
 US-09-764-870-569  
 ; Sequence 569, Application US/09764870  
 ; Patent No. US20020042386A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT214  
 ; CURRENT APPLICATION NUMBER: US/09/764,870  
 ; PRIORITY FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 646

SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 569  
LENGTH: 2332  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-870-569

## Alignment Scores:

Pred. No.:	0	Length:	2332
Score:	776.00	Matches:	776
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	26.55%	Indels:	0
DB:	9	Gaps:	0

US-09-916-849A-3 (1-2323) x US-09-764-870-569 (1-2332)

```
QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273
DB 3 ACCCAAGCTTCAAGGAGTCAAGGCGAGACCAAGGCGATGCCCGACGAAGTGCCTGGCT 62
QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlu 293
DB 63 AACCTCAACATCTTGTTGTTATGACACCAATGACATGACCTGTGTGTTGAGCAGCAGAG 122
QY 294 TyrIysGlnSerLeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValAlaAla 313
DB 123 TACAAGGAGAGCTCAGAGGAGAACTGAGAGTTGGCTATGAGTGTCTCACTGTCAAGGCGC 182
QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGly 333
DB 183 ACGAGTGTGATGAGCCCTCCCATGCAATATTCGACCGCTGTGAGAGGAGGTCTGAG 242
QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353
DB 243 GGGAGCCCTCTGAAGCTCTTGAGATCGACCTCTCTGGGGTATCCGAACCCGTGGC 302
QY 354 ProValAspArgGluGluValGlnSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373
DB 303 CCTGTGATCGGGAAGAGGTGATCTCAACGATCGATGAGGCAAGTACCAAGGT 362
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsn 393
DB 363 CGGAGCCGGGCTCTCGAGATCAACAGCCGCTGTTCTTCTGTGAGAGATACAT 422
QY 394 AspAsnAlaProGlnPheSerGluIysArgTyrValIleGlnValArgGluAspValThr 413
DB 423 GATATGCCCCCACTTATGTAAGAACCCCTATGTGCTCAGGTGAGGAGATGTGACT 482
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433
DB 483 CCAGGGGCCCCAGTACTCCGAGTCAACGCTCGATCCAGACAAAGGAGCAATCCGTG 542
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453
DB 543 GTGCCTATAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
QY 454 GlyAlaLeuAspValIleSerProLeuAspTyrGlnThrThrIleGlyTyrThrLeuArg 473
DB 603 GGAAGCTCTGAGATGTGTGAGCCCTCTTGACTATGAGACCAAGAGATGACCTTACG 662
QY 474 ValAlaGlnIleAspGlyValArgProProLeuSerAsnValSerGlyLeuValThrVal 493
DB 663 GTGCAGACAGAGAGTGTGCGCTCCCACTCTAATGTCTCGGCTGTGTGACATGTA 722
QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513
DB 723 CAGGTCTCGAATATCAAGACATGCCCCCATCTTGTCAGACACCCCTTCCAGGCTACT 782
QY 514 ValLeuGlnSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533
DB 783 GTCTCGAGAGAGGTCCCTTATGCTACCTGTCTCCATGTCTCAAGCTATGACCTCAT 842
```

```
QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValIleAspPheProPhe 553
DB 843 GCTGTGTCATATGCCCCGCTGGAATATACCTTCTGTGGGAGTGGAGCATGATCCCTTCC 902
QY 554 ThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluGluVal 573
DB 903 ACCATCAACAAATGACACAGGCTGATCTGTGTGCTGTCTGTCACTGACCGGAGAGATT 962
QY 574 AspPheTyrSerPheGlyValIleGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593
DB 963 GATTTCTACAGCTTTGGGGTGTGAAGCTCGACCAATGACATCCAGCATCACTACCTCC 1022
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnAspProThrPheThrGlnPro 613
DB 1023 GCCAGTGCAGCGTGAAGTGTCTGTGATGTCAACGACAAATTCACCTTATACCAACCA 1082
QY 614 GluTyrThrValAlaGlnLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer 633
DB 1083 GAGTACACAGTGCAGGCTCAATGAGAGATGCAAGCTGTGGACACAGCGGTGACGATGCA 1142
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653
DB 1143 GCTGTGACCTGTATGCTCATATGATGTATCATCACTACAGATCAACAGTGGCAATATCGA 1202
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673
DB 1203 AACGCTTCTTCATCACACGCAAGAGTGTGGCTGTGATCTCTTGCCTGCCACTG 1262
QY 674 AspTyrIysLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693
DB 1263 GACTCAAACTTGAAGCGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1322
QY 694 AspThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisAspProValPheGln 713
DB 1323 GACACGGCAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1382
QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrValValLeu 733
DB 1383 AGCTCCCATATACAGGAATGTTAATGAGACCGCGGCGGACACACCGTGTGTGTGT 1442
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753
DB 1443 ATCAAGCGCACAGATGAGAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1502
QY 754 SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu 773
DB 1503 AGCATCCCAAGTCCCGCATGTGATGACAGACACGGGGCTGTCAACCCAGCGTGAAGT 1562
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyTyrPro 793
DB 1563 GACTACGAAGACCAAGTGTCTTACACCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1622
QY 794 GlnIysSerAspThrTyrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro 813
DB 1623 CAGAAATCCAGACACACTTACCTGTGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1682
QY 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833
DB 1683 CAGTTCTCGAGAGCTCTTACACAGGAGAGTGTCTATGAGATGTGCCACCTTACATGAC 1742
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTyrThr 853
DB 1743 GTCTCGAGATCTCAGACATGATCGTGAATCTGTGATGTGATGTGATGTGATGTGATGTG 1802
QY 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGlnIleThrSerGlyTyrVal 873
DB 1803 TTCCAGAGAGCGACATGAGAGAGTGTGATCTTATGTGTGATGTGATGTGATGTGATGTG 1862
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla 893
DB 1863 CGAAGCGCTACGAGAGGTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 1922
QY 894 ValAspIysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913
```

Db	1923	GTGGACAAGGGGATGCCCCCGACCCCGACAACCTATGAAAGTACAGCTCATCTGTGGAT	1982
Qy	914	ValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGluAsn	933
Db	1983	GTGATATGACATCCCCCTGCTCTTGGAGAGAGATGATGTATGTGTGTGTAAGAAAGAAC	2042
Qy	934	SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn	953
Db	2043	AGCCCATTTGGGGTACGCGGTGGCCCGGGTCAACAGCATGACCCCGATGAAGGACCAAT	2102
Qy	954	AlaGlnIleMetCysArgIleLevalGluGlyAsnIleProGluValPheGlnLeuAspIle	973
Db	2103	GCCCAAGATTATGTACCAAGATTGTGAGGGCAACATCCCTGAGGCTTCCACACTGACATC	2162
Qy	974	PheSerGlyGlyLeuAlaThrAlaLeuValAspLeuAspTyrArgIleAspArgProGluTyrVal	993
Db	2163	TTCTCCGGGGAGCTGACAGCCCTGTGTACCTTGAACCTACGAGAGACCGGCTGTGATACGTC	2222
Qy	994	LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValIleValArgLeu	1013
Db	2223	CTGTGTCATCCAGGCCACGCTCAGCTTCCTGTGGTGAAGCCGGGCTACAGTCCAGTCGCGCTC	2282
Qy	1014	LeuAspArgAsnAspAsnProProValLeuGlnAsnPheGluIleLeu	1029
Db	2283	CTTATACCGCAATGACACCCACAGGTGTGGGCAACTTGTGATGCTTT	2330

Db	243	GGAGAGCCCTCTGAAGCTTTGAGATGACCTCGCTGSGGGGATCCAGACCCGGAGC	302
Qy	354	ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly	373
Db	303	CCGTGGATGGGAGAGGTGGAACTCTACCACTGACGGTAAGGCAAGTGCACAGGT	362
Qy	374	ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn	393
Db	363	CGGAGCCGGGTCTCGAGTACCAAGCGCGCTGTTTCTTTCTGTGGAGATGACAAAT	422
Qy	394	AspAsnAlaProGlnPheSerGluValArgTyrValValGlnValArgGluAspValThr	413
Db	423	GATTAATCCCCCAGTTTACTGAGAGCGCTAATGTGTTCCAGGTGAGGAGATGTACT	482
Qy	414	ProGluValProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal	433
Db	483	CCAGGGGCCCACTATCTCCAGATCAAGCCTCGGATGAGACAAAGGGACCAATCCGCTG	542
Qy	434	ValHisTyrSerIleMetSerGlyAsnAlaArgGluGlnPheTyrIleuAspAlaGlnThr	453
Db	543	GTGCACATATGACATTCATGATGGCAATGCTCGGGGACAGTTTATCTGGATGCCCAACT	602
Qy	454	GluValIleuAspValValSerProIleuAspTyrGluThrThrIleArgGluValTyrThrIleuArg	473
Db	603	GGAAGCTTGATGTGGAGACCCCTCTTGACTTAAGACGACCAAGAGATCAACCTTAGG	662

RESULT 6  
 US-10-125-540-569  
 Sequence 569, Application US/10125540  
 Publication No. US2003005875A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: Ptz1401  
 CURRENT APPLICATION NUMBER: US/10/125,540  
 CURRENT FILING DATE: 2002-04-19  
 Prior Application removed - See File Wrapper or Palm  
 NUMBER OF SEQ ID NOS: 646  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 569  
 LENGTH: 2332  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-125-540-569

Assignment Scores:		
Pred. No.:	0	Length: 2332
Score:	776.00	Matches: 776
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	26.55%	Indels: 0
DB:	15	Gaps: 0

US-09-916-849A-3 (1-2923) X US-10-125-540-569 (1-2332)

QY	254	ThrHisValPheArgValThrIaIaGnaPheIsgIwMeProArgArgSerLalaLeuAla	273
Db	3	ACCCAGCTTCAGGGGTACGGGACGAGACACAGGATGCCCGACGAAGTCCCTGGCT	62
QY	274	ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheIugIuIngIuIu	293
Db	63	ACACTCACCATCTTGTTACTGACACCAATGACCAATGACCTGTGTGTCAGCAGCAGGAG	122
QY	294	TyrIlygIuSerIeuArgIuasnIeuGluValGlyTyrGluValLeuThrValArgAla	313
Db	123	TACAAAGAGAGGCTCAGGAGGAACTCGAGAGTGGCTTAAGAGGTCTCACTGCACAGGGCC	182
QY	314	ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgIeuIeuGluIsgIySerGly	333
Db	183	ACGAGTGGGATGCCCTCCCAATGCCAATATTCTTACCGCTGCTGGAGGGGTCTGGG	242
QY	334	GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly	353

Db	243	GGAGAGCCCTCTGAAGCTTTGAGATGACCTCGCTGSGGGGAGATCCGACCCCGTGGC	302
Qy	354	ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly	373
Db	303	CCGTGTGATGGGAGAGAGGTGAGATCCTACACACTGACGGTAAAGGACAGTACAGAGGT	362
Qy	374	ArgAspProGlyProAspSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn	393
Db	363	CGGAGCCCGGGTCTCTCGAGTACACAGCCGCTGTTTCTTCTCTGTGGAGGATGACAT	422
Qy	394	AspAsnAlaProGlnPheSerGluValAspTyrValValGlnValArgGluAspValThr	413
Db	423	GATTAATCCCCCGAGTTTAACTGAGAAAGCGCTAATGTGCTCAGGTGAGGGAGATGTGACT	482
Qy	414	ProGluValaProValLeuArgValThrAlaSerAspArgAspGlySerAspAlaVal	433
Db	483	CCAGGGCCCCAGTACTCCAGTGCACAGCTCCGATCGAACAAGGGAGACATGCGCTG	542
Qy	434	ValHisTyrSerLileMetSerGlyAspAlaArgGlyGlnPheTyrLeuAspAlaGlnThr	453
Db	543	GTGCACATATGACATCAATGATGGCAATGCTCGGGGACAGTTTATCTGATGCCCAACT	602
Qy	454	GlyValaLeuAspValValSerProLeuAspTyrGluThrThrTyrGluGlyTyrThrLeuArg	473
Db	603	GGAGCTCTGATGTGGTGAAGCCCTCTTGACTTAAGACGACCAAGAGATACACCTTAAG	662
Qy	474	ValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrVal	493
Db	663	GTGGAGACACAGATAGTGTGGCCCTCCCACTCTCTAATGTCTGTGGCTTGATGACAGTA	722
Qy	494	GlnValLeuAspLileAsnAspAsnAlaProLilePheValSerThrProPheGlnAlaThr	513
Db	723	CAGGTCTGTGATATTCACACGACAAATGCCCACTTCTGTGACGACCCCTTTCCAGGCACT	782
Qy	514	ValLeuGluSerValProLeuGlyTyrTyrLeuValLeuHisValGlnAlaLileAspAlaAsp	533
Db	783	GTCTGTGAGAGCGTCCCTTAAGCTACCTGTGTTCTCATGTCCAGGCTTAACGACGTGAT	842
Qy	534	AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe	553
Db	843	GCTGTGACATAGCCCGCTCGAATACCGCTTGCTGGGTGGGACATGACTTCCCTTC	902
Qy	554	ThrLileAsnGlnGlyThrGlyTyrPrlLileSerValAlaAlaGluLeuAspArgGluVal	573
Db	903	ACCATCAACATGGCACAGGCTGATTCGTGTGCTGTAACGTGACCGGAGAAAGTT	962
Qy	574	AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer	593
Db	963	GATTTCTACAGCTTTGGGGTAGAAGCTCGAGACCATGGACCTCCAGCACTCAGCGCTCG	1022
Qy	594	AlaSerValSerValThrValLeuAspValAsnAspAsnAspProThrPheThrGlnPro	613
Db	1023	GCCAGTGTCAAGCGGTGCTGTGAGAGTCAACACAAATCCAACTTATCCCAACCA	1082
Qy	614	GluTyrThrValaArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer	633
Db	1083	GAGTACACAGTGGCGGCTCAATGAGAGTGCAGCTGTGGCACACGCGTGTGACGCTGTCA	1142
Qy	634	AlaValaAspArgAspAlaHisSerValLileThrTyrGlnLileThrSerGlyAsnThrArg	653
Db	1143	GCTGTGACCGGTATGCTCATATGTCATACCTACACAGATCACACAGTGGCAATATCTGA	1202
Qy	654	AsnArgPheSerLileThrSerGlnSerGlyGlyLeuValSerLileuAlaLeuProLeu	673
Db	1203	AACCGCTTCTCATACACAGCCAAAGTGTGGTGGGCTGTGATCTCTTGCCCTCCACTG	1262
Qy	674	AspTyrLysLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln	693
Db	1263	GACTACAAACTTGAAGCGGACGATATGTGTGGCTGTACCGCTCCGATGGCACTCGGCAG	1322
Qy	694	AspThrAlaGlnLileValValaAsnValThrAspAlaAsnThrHisArgProValPheGln	713

Db 1323 GACACGGCACAAGATTGCTGATGTCACCGACGCCAACCCATGCTCTGCTTCAG 1382  
 Qy 714 SerSerHisrThrValaenValaenGluaspArgProAlaGlyThrThrValaValaLeu 733  
 Db 1383 AGCTCCCACTATACAGTAATGTTAATGAGACCGGGCCGACGACCAACGAGTGCTG 1442  
 Qy 734 IleSerAlaThrAspGluuAspThrGlyGluAsnAlaArgGlyLeuThrThrPheMetGluAsp 753  
 Db 1443 ATCAGGGCCACGAGATGAGACACAGGTGAGATGCCGACATCACTTCATGAGGAGAC 1502  
 Qy 754 SerIleProGlnPheArgIleAspAlaAspThrGlyValaValaThrThrGlnAlaGluLeu 773  
 Db 1503 AGCATCCCGCCAGTTCGCCATCGATCGACGACACCGGGGCTCTCACACCCAGCTGAGCTG 1562  
 Qy 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
 Db 1563 GACTACAGAGACCAAGTGTCTTACACCTGCGCATTACTCTCGGGACATGSCATTTCCC 1622  
 Qy 794 GlnIleSerAspThrThrTyrIleuGluIleLeuValaAsnAspValaAsnAspAsnAlaPro 813  
 Db 1623 CAGAAAGTCCGACACCACTACCTGAGAGATCTGGTGAACGATGAAATGCAATGCCCCCT 1682  
 Qy 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValaProPheThrSer 833  
 Db 1683 CAGTTCCTGGAGACTCTTACAGGGCAGTGTCTTAAGAGATGTGCCACCTTCATCACTAGC 1742  
 Qy 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTyrThr 853  
 Db 1743 GTCTCTCAGATCTCAGCACTGATCGTGAATTCGACTTAATGCGAGGGCTTCTTCAAC 1802  
 Qy 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGlySerThrSerGlyIleVal 873  
 Db 1803 TTCCAAGAGGACCAATGAGACGAGTGAATTTATGTTAGTCAACGTCGAGGCATCGTG 1862  
 Qy 874 ArgThrLeuArgThrLeuAspArgGluAsnValaIleGlyTyrValaLeuArgAlaTyrAla 893  
 Db 1863 CGAAAGCTACGAGAGCTGATCGAGAGAACTGGCCCAATATCTTGGGGCATATGCA 1922  
 Qy 894 ValAspLeuGlyMetProProAlaArgThrPrometGluValaThrValaLeuAsp 913  
 Db 1923 GTGCAAGAGGGAGTCCCGCAGCCCGCACCTATGGAAGTGAAGTCACTGCTGTGGAT 1982  
 Qy 914 ValAspAspAsnProProValaPheGluGlnAspGluPheAspValaPheValaGluGluAsn 933  
 Db 1983 GTGAATGACATCTCCCTGCTTGTAGCAGAGATGATGATTTGATGTTGTGAGAAAGAAC 2042  
 Qy 934 SerProIleGlyLeuAlaValaIleArgValaThrAlaThrAspProAspGluGlyThrAsn 953  
 Db 2043 AGCCCATTTGGGCTACCGTGGCCGGGTCAACAGCACTGACCCCGATGAAGGCACCAAT 2102  
 Qy 954 AlaGlnIleMetTyrGlnIleValaGluGlyAsnIleProGluValaPheGlnLeuAspIle 973  
 Db 2103 GCCCAATATTATGACAGATTGTGGAGGGCAACATCCCTGAGGGTCTTCCAGCTGAGCANTC 2162  
 Qy 974 PheSerGlyGluLeuThrAlaLeuValaAspLeuAspTyrGluAspArgProGluTyrVala 993  
 Db 2163 TTCTCCGGGAGGTGACAGCCCTGTGAGCTTAAGCTACGAGGACCGGCTGAGTACGTC 2222  
 Qy 994 LeuValaIleGlnAlaThrSerAlaProLeuValaSerArgAlaThrValaIleValaArgLeu 1013  
 Db 2223 CTGGTATTCAGGCCACGTCAGCTCTCTGGTGAAGCCGGGCTCAAGTCAAGTCCGCTC 2282  
 Qy 1014 LeuAspArgAsnAspAsnProProValaLeuGlyAsnPheGluIleLeu 1029  
 Db 2283 CTGTAACCGAATGACAACCAACGAGTGGGCACTTGAAGATCCTT 2330  
 RESULT 7  
 US-10-176-847-99  
 ; Sequence 99, Application US/10176847  
 ; Publication No. US20030068636A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Velby, Peter Ole  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST  
 ; FILE OF INVENTION: AND OVARIAN CANCER  
 ; FILE REFERENCE: MRI-039  
 ; CURRENT APPLICATION NUMBER: US/10/176,847  
 ; CURRENT FILING DATE: 2002-06-21  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 99  
 ; LENGTH: 2391  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-176-847-99  
 Alignment Scores:  
 Pred. No.: 0 Length: 2391  
 Score: 729.00 Matches: 729  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 24.94% Indels: 0  
 DB: 15 Gaps: 0  
 US-09-916-849A-3 (1-22923) x US-10-176-847-99 (1-2391)  
 Qy 2186 ArgTyrGluAlaLeuArgGlyGluGlnProProAspLeuGluThrThrValaIleLeuPro 2205  
 Db 19 CGCTACGAGGCCCTCGTGGGAGCAGCCCGCGACCTTGACACAAAGTCAATCTGCT 78  
 Qy 2206 GluSerValaPheArgGluThrProProValaValaArgProAlaGlyProGlyGluAlaGln 2225  
 Db 79 GAGTGTCTTCAAGAGAGAGCCCGCCCGTGTGACGCGCCGAGGCGCCGAGAGAGCCGAC 138  
 Qy 2226 GluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGlyGluAla 2245  
 Db 139 GAGCCAGAGAGAGCTGCGACCGCGACAGCGACCGGACCTGACGAGCTGACGAGGCT 198  
 Qy 2246 ValAlaSerValaIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspPro 2265  
 Db 199 GTGGCAGCGTCATATCTACCGCACCTGGCCGGGCTTACCTGCTCATATATGACCT 258  
 Qy 2266 AspIleAspSerLeuArgValaProAspArgProIleIleAsnThrProValaValaSerIle 2285  
 Db 259 GACAAAGCCAGCTTACAGATGCCCAACGCCCGATCAACACACCCGTGTGACATC 318  
 Qy 2286 SerValaHisAspAspGluGluLeuLeuProAlaLeuAspLysProValaThrValaGln 2305  
 Db 319 AGCGTCCATGATGATGAGAGAGCTTCTGCGGGCCCTGACAAACCGTCAACGCTGACG 378  
 Qy 2306 PheArgLeuLeuGluThrGluGluArgThrTyrProIleCysValaPheTyrAsnHisSer 2325  
 Db 379 TTCCGCTGTGAGACAGAGAGAGGACCAAGCCCATCTGTGCTTCTGGAACCATTCAC 438  
 Qy 2326 IleLeuValaSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValaValaPheArgAsn 2345  
 Db 439 ATCTGGTCACTGGGACAGAGTGGCTGTGGCCAGAGCTGTGAAGTGTCTTCCGCAAT 498  
 Qy 2346 GluSerHisValaSerCysGlnCysAsnHisMetThrSerPheAlaValaLeuMetAspVala 2365  
 Db 499 GAGAGCCACGTCAGCTGCGACATGCAACCAATGACAGACTTCGCTGTCTCATGACGCT 558  
 Qy 2366 SerArgArgGluAsnGlyGluIleLeuProLeuLysThrIleuThrTyrValaAlaLeuGly 2385  
 Db 559 TCTCGCGGAGAAATGGGAGATCTCTGCACTGAAGACATGACATGATGAGCTTAGGT 618  
 Qy 2386 ValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArgIleLeuArg 2405  
 Db 619 GTCACTTGGCTGCTCTTCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 678  
 Qy 2406 SerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuVala 2425  
 Db 679 TCACAACCAACGAGCATCCGACATCACTGACAGCTGCGGCGCTGCTGCTGCTGCTGCTG 738  
 Qy 2426 PheLeuGluGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValaIleAlaIleLeu 2445



```

Db      739 TTCCTCTGGGAATCAACGAGGCTGACCTTCCTTTGCTGCACAGTCACTATGCTCATCTG 798
Qy      2446 LeuH1ePheLeuTyrLeuCyThrPheSerTrpAlaLeuLeuGluValaLeuH1sLeuTyr 2465
Db      799 CTGCACCTTCTGTACCTCTGACACCTTTTCTGGGGCTCTGCTGGAGGCTTTGACCTGTAC 858
Qy      2466 ArgAlaLeuThrGluValaArgAspValaAsnThrGlyProMetArgPheTyrTyrMetLeu 2485
Db      859 CGGGCACTCACTAGAGGCGCGATGTCACACCGGGCCCACTGCGCTTCACTACATGCTG 918
Qy      2486 GLYTTPGlyValaProAlaPheIleThrGlyLeuAlaValaGlyLeuAspProGluGlyTyr 2505
Db      919 GGGTGGGGCGTGGCTTCCCTTCATCAAGGGGCTAGCGGTGGGCTTGAACCCCGAGGGCTAC 978
Qy      2506 GlyAsnProAspPheCyThrLeuSerIleTyrAspThrLeuIleThrPheAlaGly 2525
Db      979 GGGAACTCTGACCTTCTGCTGGCTCTCATATGACACGCTCATCTGGAGTTTGTGGC 1038
Qy      2526 ProValaAlaPheAlaValaSerMetSerValaPheLeuTyrIleLeuAlaAlaArgAlaSer 2545
Db      1039 CCGGTGGCTTTGCTCCCTCTGATGAGTGTCTTCTGTACATCTGGCGGGCCGCTCC 1098
Qy      2546 CyAlaAlaGlyArgGlnGlyPheGlyValaGlyValaGlyProValaSerGlyLeuGlnProSer 2565
Db      1099 TGTGTGCCCCAGCGGCGAGGCTTTGAGAAAGAGTCTGTCTGTGGGCTTGCAGCCCTCC 1158
Qy      2566 PheAlaValaLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuSerValaAsnSer 2585
Db      1159 TTGCGGCTCTCTGCTGCTGACCGCCACCTGGCTGCTGCTCTCTGTCAACAGC 1218
Qy      2586 AspThrLeuLeuPheH1sTyrLeuPheAlaThrCyAsnCyAlaGlnGlyProPheIle 2605
Db      1219 GACACCTCTCTTCCACATCACTCTTGTGCTACCTGCAATTGCAATCAAGGGCCCTTCATC 1278
Qy      2606 PheLeuSerTyrAlaValaLeuSerGlyValaArgValaAlaLeuYsLeuAlaCySer 2625
Db      1279 TTCTCTCTCATATGTGTGTACAGAGGTCCGAAAGCACTCAAGCTTGTCTGACG 1338
Qy      2626 ArgLysProSerProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsn 2645
Db      1339 CGCAAGCCGAGCCCTGACCTGTGCTGACCAACCAAGCCCTGACCTGTCTTCAAC 1398
Qy      2646 CyPProSerProTyrAlaAspGlyValaGlyTyrGlnProTyrGlyAspSerAlaGlySer 2665
Db      1399 TGCCCCAGGCCCTTACGAGATGGCGGCTGTACCAAGCCCTTACGAGAACTCGCGGCTCT 1458
Qy      2666 LeuH1sSerThrSerArgSerGlyLysSerGlnProSerTyrIleProPheLeuAlaGly 2685
Db      1459 CTGCAAGCAACAGTGTGCTCGGGCAAGATCAAGCCCAAGCTCAATCCCTTGTGTGAGG 1518
Qy      2686 GlnGlySerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeu 2705
Db      1519 GAGGAGTCCGCACTGAACCTTGCCCAAGGGCCCTTGCGCTGGGGATCCAGGCAACCTG 1578
Qy      2706 PheLeuGlnGlyGlnAspGlnGlnH1sAspProAspThrAspSerAspSerAspLeuSer 2725
Db      1579 TTCTCTGAAAGTCAAGACACGAGCATGATCTGACACGCACTCCCAAGTGCCTGTCC 1638
Qy      2726 LeuGlnAspAspGlnSerGlySerTyrAlaSerThrH1sSerSerAspSerGlyGlnGln 2745
Db      1639 TTGAGAGACCAACAGATGTGCTCTATGCTCTTACCACTCAACAGACAGTGGAGAGAA 1698
Qy      2746 GlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeu 2765
Db      1699 GAAAGAGAGAGAGAGAGAGAGCGGCTTCTCGAGAGAGAGGCTGGATACCTGTCTG 1758
Qy      2766 GlyProGlyValaGlyValaGlyLeuPheLeuH1sSerThrProLysAspGlyGlyPro 2785
Db      1759 GGGCTGTGAGACAGAGACAGTCCCTGTGCAAGTACCTCCCAAGATGGGGCCCAAGGACCT 1818
Qy      2786 GlyValaAlaProTyrProGlyLysAspPheGlyThrThrAlaLysGlyLysSerGlyLysGly 2805
Db      1819 GGCAGAGCCCTCTGGCCAGAGACTTTGGGACCAACAGCAAAAGAGTATGGCAACGGG 1878

```

```

Qy      2806 AlaProGlnGlnArgLeuArgLysAspAlaLeuSerArgGlnGlySerLeuGly 2825
Db      1879 GCGCTGAGAGAGCGGCTCGGAGAGATGGAATGCCCTGTCTCAAGAGGGGCTCTAAGG 1938
Qy      2826 ProLeuProGlySerSerAlaGlnProH1sLysGlyIleLeuValaLysCyLeuPro 2845
Db      1939 CCCCTTCAAGGCTCTTGTGCCAGCCTCAAAAGCATCTTAAGAAAGATGTGTGCC 1998
Qy      2846 ThrLysSerGlyLysSerSerLeuLeuArgProLeuGlnGlnCyThrGlySerSer 2865
Db      1999 ACCATACGAGAGAGAGAGAGAGCTCTGCGGCTCCCTGAGCAATGCAAGGGTCTTCC 2058
Qy      2866 ArgGlySerSerAlaSerGlnGlySerArgGlyGlyProProProArgProProArg 2885
Db      2059 CGGGCTCTCGGCTAGTGAAGGAGGAGCGGGGGGCGGCCCTCCCGCCACCGCCCGG 2118
Qy      2886 GlnSerLeuGlnGlnGlnLeuAsnGlyValaMetProIleAlaMetSerIleValaGly 2905
Db      2119 CAGAGCTTCCAGAGACACTGAACGGGGTCAATGCCCATCGCATAGATCAAGCAGG 2178
Qy      2906 ThrValaAspGlyAspSerSerGlySer 2914
Db      2179 ACGGTGATGAGAGACTTCGTACAGGCTCC 2205

RESULT 8
US-09-764-870-570/C
; Sequence 570, Application US/09764870
; Patent No. US0020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 570
; LENGTH: 2077
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-570

Alignment Scores:
Pred. No.: 0 Length: 2077
Score: 691.00 Matches: 691
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.64% Indels: 0
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x US-09-764-870-570 (1-2077)
Qy      2254 ThrH1sValaPheArgValaThrAlaGlnAspH1sGlyMetProArgArgSerAlaLeuAla 273
Db      2075 ACCCAAGCTTCAAGGAGTCAAGCGGAGAGACCAAGGCAAGGCCCGCAAGTGCCTGTGCT 2016
Qy      2274 ThrLeuThrIleLeuValaThrAspThrAsnAspH1sAspProValaPheGlnGlnGlnGln 293
Db      2015 AACCTCAACATCTTGTGTATGACACCAATGACATGACACCTGTGTGTGAGCAGAGAG 1956
Qy      2294 TyrLysGlnSerLeuArgLysAsnLeuGlnValaGlyTyrGlnValaLeuThrValaArgAla 313
Db      1955 TACAAAGAGAGCTCAAGGAGAGAACTGAGAGGTGTGATGAGTGTCTCATGTCTGAGGCG 1896
Qy      314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGly 333
Db      1895 ACGAGTGTGAGTCCCTCCCAATGCAATGCAATTTCTGTACCGGCTCTGAGAGGAGTCTGAG 1836
Qy      334 GlySerProSerGlnValaPheGlnLysAspProArgSerGlyValaIleArgThrArgGly 353
Db      1835 GGCAGCCCTCTGAAGCTTTGAGATGACCTTCCTCTGGGGTATCTCGAACCCTGTGCTG 1776

```

```

QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluIleAspArgGlnGly 373
DB 1775 CCGTGGATCGGGAAGAGGTGGAATCTCAACAGCTGACGGTAGAGCAAGTGCACGGGT 1716
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393
DB 1715 CCGAGACCCGGGCTCTCGAGGTACCAAGCCGCTGTTCTCTGTGTGAGATGACAT 1656
QY 394 AspAsnAlaProGlnPheSerGlnGlySerArgTyrValGlnValArgGluAspValThr 413
DB 1655 GATTAATGCCCCCAAGTTAGTAGAGCCCTATGTGTCTCAGGTGAGGAGATGTGACT 1596
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433
DB 1595 CCAGGGGCCCCAGTACTCCGAGTACAGAGCTCGAGTCAGAGCAAGGGGAGCAATGCCGTG 1536
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453
DB 1535 GTGCACTATAGCATCATAGTGGCAATGCTCGGGGACAGTTTATCTGTGATGCCCAACT 1476
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGlnTyrThrLeuArg 473
DB 1475 GGAAGCTCTGAGTGTGTGAGCCCTCTTGACTATGAGACCAAGAGATACACCTTACGG 1416
QY 474 ValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493
DB 1415 GTGCCAGCAACAGATGTGGCCGCTGCCCACTCTAATGTCTGTGGCTTGTGACAGTA 1356
QY 494 GlnValLeuAspAlaAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513
DB 1355 CAGGTCTCGAGTATCAACCAATGCCCCCATCTTGTGAGCAACCCCTTTCAGAGCTACT 1296
QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533
DB 1295 GTCTCGAGAGAGTGTCCCTTAAGCTTACCTGTCTCCATGTCAGAGCTATCAAGCTGAT 1236
QY 534 AlaGlyAspAsnAlaArgLeuGlnTyrArgLeuAlaGlyValGlyHisAspPheProPhe 553
DB 1235 GCTGTGTGCAATGCCCGCTGGAATPCCGCTTGTGGGGTGGGACATGACTTCCCTTTC 1176
QY 554 ThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluGlnVal 573
DB 1175 ACCATCAACATAGGACCAAGCTGAGTCTGTGGCTGTGAACCTGAGACCGGAGAGAGTT 1116
QY 574 AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593
DB 1115 GATTTCTACAGCTTGGGGTAGAGCTCGAGACCAATGGCACTCCAGCACTCATGCTCTCG 1056
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnPro 613
DB 1055 GCCAGTGTACAGCGTCACTGTCTCGATGTCAACGACCAACATCCAACTTACCCACCA 996
QY 614 GlnTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer 633
DB 995 GAGTACACAGTCCGCTCATATAGATGAGTGTGGGACCAAGGCTGTGACCGGTGCA 936
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653
DB 935 GCTGTGAGACCGTGTCTCATGTGTCACTCACTTCCAGATACCAAGTGGCAATATCTGA 876
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673
DB 875 AACCCCTCTCCATACCAAGCAAGTGTGGTGGCTGTGATCCCTTGCCTGCACTG 816
QY 674 AspTyrValLeuGluAspGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693
DB 815 GACTACAACTTGTAGGGGAGATGTGTGGCTGTATCCGCTCGAGTGGCACTGGGAG 756
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713
DB 755 GACACGGGACCAATTTGTGTGATGTCAACGACGCGCAACCACTGCTCTGTCTTTCAG 696

```

```

QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu 733
DB 695 AGCTCCCACTATACAGTAAATGTTTATAGAGACCGCGCGGACGAGCAACAGGTGTGCTG 636
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753
DB 635 ATCAAGCGCACGAGAGAGAGACAGGTGAGATGCCCACTCACTTCAATGAGAGAC 576
QY 754 SerIleProGlnPheArgIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGlnLeu 773
DB 575 AGCATCCCCCAAGTTCGATGATGATGACACAGGGGGCTGTGCACCAAGGCTAGAGCTG 516
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793
DB 515 GACTATGAAGACCAAGGTCTTACACCTTGGCAATTACTGTCTGGGACATATGCACTTCCC 456
QY 794 GlnTyrSerAspThrThrTyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaPro 813
DB 455 CAGAAGTCCGACACCACTTACCTGTGAGATCCTGTGAAACGATGAATGACATGCCCT 396
QY 814 GlnPheLeuAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833
DB 395 CAGTTCCTCGAGACTCTTACCAAGGCAGTGTCTATGAGATGTGCCACCTTCACTAGC 336
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853
DB 335 GTCTCGACAGTCTCAGCCACTGATGTGATCTGGAATTAATGGAGGGTCTTTCACACC 276
QY 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal 873
DB 275 TTCAGAGAGCGACAGATGAGAGAGGTGACTTTATGTTGATGTCACCTCAGGACATCGTG 216
QY 874 ArgThrLeuAspArgLeuAspArgGluAsnValAlaGlnTyrValLeuAspAlaTyrAla 893
DB 215 CGAAGCTCGGAGAGCTGATCGAGAGACCTGGCCCACTATGTCTTTCGGGCAATATCA 156
QY 894 ValAspGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913
DB 155 GTGACAAAGGGAGATGCCCGCCAGCCGCAACCTATGGAAGTACAGTCACTGTGTGAT 96
QY 914 ValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGlnAsn 933
DB 95 GTGAATGCAATCCCCCTGTCTTGTAGCAGATGATGATTTGATGTTGTGAAAGAAC 36
QY 934 SerProIleGlyLeuAlaValAlaArgValThr 944
DB 35 AGCCCAATGGGCTAGCCGTGGCCCGGCTACA 3

```

RESULT 9  
 US-10-125-540-570/c  
 ; Sequence 570, Application US/1012540  
 ; Publication No. US2003059875A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT214C1  
 ; CURRENT APPLICATION NUMBER: US/10/125,540  
 ; PRIORITY FILING DATE: 2002-04-19  
 ; Prior Application removed - See file wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 646  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 570  
 ; LENGTH: 2077  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-125-540-570

Alignment Scores:  
 Pred. No.: 0  
 Score: 691.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 23.64%

Length: 2077  
 Matches: 691  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0

DB: 15 Gaps: 0  
US-09-916-849a-3 (1-2923) x US-10-125-540-570 (1-2077)  
QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProAlaArgSerAlaLeuAla 273  
Db 2075 ACCCAGGCTCTTCAGGGGCGACGGCGCAGAGCACGGCGATGCCCGACCAATGCGCTGGCT 2016  
QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlu 293  
Db 2015 ACCTACCACTCTGGTTACTGACCAACCAATGACCAATGACCTGTGTTCAGACAGAGAG 1956  
QY 294 TyrIleGluSerLeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
Db 1955 TACAAAGAGAGCCCTCAGGAGAGAACTTGAGGTGGCTATAGAGGTCCTCACTGCAAGGCC 1896  
QY 314 ThrAspGlyAspAlaProProAlaAsnIleLeuTyrArgLeuLeuGluGlySerGly 333  
Db 1895 ACCGATGTGATCCCTCCCAATGCCAAATATCTGTATCCGCTGCGAGGGGTCTGGG 1836  
QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
Db 1835 GGGAGCCCTCTGAAGCTTTGAGATGACCTCGCTCGGGTGATCCAAACCCGTGAC 1776  
QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
Db 1775 CCGTGGATCGGGAAGAGGTGGAATCTTACAGCTGACGGTACGAGGCAATGACAGGGT 1716  
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393  
Db 1715 CGGAGCCCGGGTCTCGAGATACCAAGCCGCTGTTTCTTCTGCGAGGATGACAAAT 1656  
QY 394 AspAsnAlaProGlnPheSerGluLysArgTyrValValGlnValArgGluAspValThr 413  
Db 1655 GATTAATCCCTCCAGATTAGTGAGAGACCTATATGTCCAGGTGAGGAGAGATGATCT 1596  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
Db 1595 CCAGGGGCCCCAGTACTCCAGTCAACAGCTCGGATCGAGCAAGGGGAGCAATGCGGTG 1536  
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
Db 1535 GTGCACATATGACATGATGATGCGATGCTCGGGGAGAGTTTATCTGATGATGCCAGACT 1476  
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrIleArgGluTyrThrLeuArg 473  
Db 1475 GAGGCTCTGATGTGTGAGACCTCTTGAATGAGAGCAACAGAGATCACCCCTACGG 1416  
QY 474 ValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
Db 1415 GTCCGAGCAAGGATGTGGCCGCTCCCACTCTCTAATGTCTCTGGCTGGTGAACGTA 1356  
QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
Db 1355 CAGGCTCTGATATCAACAGCAATGCCCCCACTCTGTGAGCAACCTTTCACAGGCTACT 1296  
QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
Db 1295 GTCTGAGAGAGTGCCTCTTAAGCTACTGTGTCTTCCATGTCAGGCTATCGAGCTGAT 1236  
QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe 553  
Db 1235 GCTGTGTACATGCCCCCTGAGATACCGCTTGTGTGGGTGGGACATGATCTCCCTTC 1176  
QY 554 ThrIleAsnAsnGlyThrGlyTyrPheSerValAlaIleGluLeuAspAspArgGluVal 573  
Db 1175 ACCATCAACATGACCAAGCTGAGATCTCTGTGCTGTCTGACTGACCCGAGAGAGATT 1116  
QY 574 AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
Db 1115 GATTTCTACAGCTTTGGGGTGAAGCTCGAGACATGAGCACTCCAGACTCACTGCTCG 1056  
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613

Db 1055 GCCAGTGCAGCGTACGTCTCTCGATGTCAACGACCAACATTCACCTTACCCACCA 996  
QY 614 GlnTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValThrValSer 633  
Db 995 GAGTACACAGTCCGCTCAATGAGATGCAAGCTGTGGGACCAAGCTGTGACGGTGA 936  
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653  
Db 935 GCTGTGACCGGTATGCTCATAGTGTCACTACCTTACAGATCAACAGTGGCAATCTCA 876  
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673  
Db 875 AACCGCTTCCATCACCAAGAGTGGTGGCTGTATCCCTTCCCTGCGACATG 816  
QY 674 AspTyrIleLeuGlnAspGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
Db 815 GACTACAACTTGAGCGGAGTATGTGTGGCTGTATCCGCTCCGATGGCACTCGGACG 756  
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713  
Db 755 GACACGGCACAGATTTGTGTGATGTCAACCGACGCCAACCCATGTCTGTCTTTCAG 696  
QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu 733  
Db 695 AGCTCCCACTATACAGTAAATGTATATGAGACCGCGCGGACAGCAACGGTGTGCTG 636  
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753  
Db 635 ATCAGCGCCACGATGAGAGACACAGTGAATGCCCGCATCACCTCACTTCAGAGGAC 576  
QY 754 SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu 773  
Db 575 ACATGCCCACTTCCGATTCGATGATGCAAGACAGGGGGGCTGTACACCAAGCTGAGCTG 516  
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
Db 515 GACTATGAAGACCAAGTGTCTTACACCTCGGCACTTACGTCTGGGACATGCGCATTTCCC 456  
QY 794 GlnLysSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro 813  
Db 455 CAGAGTCCGACACCACTACTGAGATCTGTGTGAGACAGATGATATACATATGCCCT 396  
QY 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833  
Db 395 CAGTTCTCTGAGAGCTCTTACACAGGGCAGTGTCTATGAGATGTGCCACCTTCACTACG 336  
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTyrThr 853  
Db 335 GTCTGTGACATCTCAGCCACTGATCGTGAATCTGAGACTTAAATGCGAGGGTCTTCAACCC 276  
QY 854 PheGlnGlyIleAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal 873  
Db 275 TTCAGAGAGGCGCAGATGAGACCGGTACTTTATTTGTGATGTCACGTCAAGGATCGTG 216  
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaIleVal 893  
Db 215 CGAACGCTACGAGAGGTGATCGAGAGACGTGGCCCAATGATGTCTTGGGCGCATGTGCA 156  
QY 894 ValAspIleGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913  
Db 155 GTGAGACAGGGAGTGGCCCCCAGCCGACACCTATGAGAGTACATGACTGCTGTGTGAT 96  
QY 914 ValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGluAsn 933  
Db 95 GTGAATGACAATCCCTGCTCTTGTAGCAGAGATGATGTGTTGTGTGGAAGAGAAC 36  
QY 934 SerProIleGlyLeuAlaValAlaArgValThr 944  
Db 35 AGCCCATTTGGGCTAGCCGTGGCCCGGTGCACA 3  
RESULT 10  
US-09-843-856-1

Sequence 1, Application US/09843856  
 Patent No. US20020034785A1  
 GENERAL INFORMATION:  
 APPLICANT: SOPPET, DANIEL R.  
 LI, XI  
 RUBEN, STEVEN M.  
 TITLE OF INVENTION: CALCITONIN RECEPTOR  
 NUMBER OF SEQUENCES: 26  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
 STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
 CITY: WASHINGTON  
 STATE: D.C.  
 COUNTRY: US  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/843,856  
 FILING DATE: 30-Apr-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/970,758  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: STEPPER, ERIC K.  
 REGISTRATION NUMBER: 36,688  
 REFERENCE/DOCKET NUMBER: 1488.0660001/EKS/KMT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2600  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2603 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 652..2355  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-843-856-1  
 Alignment Scores:  
 Pred. No.: 0 Length: 2603  
 Score: 537.00 Matches: 723  
 Percent Similarity: 99.45% Conservative: 0  
 Best Local Similarity: 99.45% Mismatches: 2  
 Query Match: 18.37% Indels: 4  
 DB: 9 Gaps: 0  
 US-09-916-849A-3 (1-2923) x US-09-843-856-1 (1-2603)  
 QY 2197 AADPLEUGLThrThrValIleuProgluSerValPheArgGluThrProProValVal 2216  
 DB 175 GACCTTGAAACACATCTATCTGCTGAGTCTGTCTTCAAGAGAGCGCCCCCGTGC 234  
 QY 2217 ATGProAlaGlyProGlyValuAglngluProgluGluIleuAlaArgArgGlnArg 2236  
 DB 235 AAGCGCCGAGGCCCCGAGAGGCGCCAGAGCCAGAGACTGGCACGCGGACAGGACCG 294  
 QY 2237 HAsProGluLeuSerGlnGlyValuAvalAlaSerValIleIleTyraArgThrLeuAla 2256  
 DB 295 CACCCGAGACTGACGAGGAGTGTGAGCTGTGCGCAGGCTCATCTACCGCACCTGTGGCC 354  
 QY 2257 GtLeuLeuProHlaSerTyraProAspArgSerLeuArgValProIysArgPro 2276  
 DB 355 GGGCTACTGCTCTACTATGACCTTGACAGCGGAGCTTGAAGATCCCAAGGCGCG 414

QY 2277 IleIleAsnThrProValValSerIleSerValHlaAspAspGluGluLeuLeuProArg 2296  
 DB 415 ATCATCAACACACCCGCTGTGAGCATCACGCTTCATGATGAGAGACTTGTGCCCGG 474  
 QY 2297 AlaleuAspLysProValThrValGlnPheArgLeuLeuGluThrGluIuArgThrIlys 2316  
 DB 475 GCCGTGACAAACCCGTCACGCTGACGTTCCGCTGTGAGACAGAGAGCGGACCAAG 534  
 QY 2317 ProIleCyValaPheTrpAsnHlaSerIleLeuValSerGlyThrGlyTyrrSerAla 2336  
 DB 535 CCATCTGTGCTTTTGAAACATTCATCTCTGTCAGTGGACAGGTGGCTGTGGCGCC 594  
 QY 2337 ArgGlyCyGluValValPheArgAsnGluSerHlaValSerCyGlnCysAsnHlaWe 2356  
 DB 595 AAGAGCTGTGAAGTGTCTTCGCAATAGAACCAACGTCAGTCCAGTT-CACCACT 653  
 QY 2356 ThrSerPheAlaValIleuMetAspValSerArgArgGluAsnGlyIuIleuProle 2376  
 DB 654 GACGAGCTTCGCTGTGCTCATGAGCGTTTCTCGCGGAGAAATGGGAGATCCTGCCACT 713  
 QY 2376 WlyThrLeuThrTyrrValAlaLeuGlyValThrLeuAlaIleuLeuLeuThrPheP 2396  
 DB 714 GAAGACACTGACATACGTGCTAGGTGTGCGCCTTGCGTCCCTTGTGCTCACCTTCT 772  
 QY 2396 hPheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHlaGlyIleArgArgAsnLeuT 2416  
 DB 773 TCTTCTCATCTCTTGTGGTATCTCGGCTCCACCAACAGCGCATCCGACGTAACTGGA 832  
 QY 2416 hAlaAlaLeuGlyLeuAlaGluLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuP 2436  
 DB 833 CAGTGCCTCGGCGCTGCTGCTGAGCTGTCTTCTCTCGGAAATCAACAGGCTGACCTCC 892  
 QY 2436 RobPheAlaCyThrValIleAlaIleuLeuHlaPheLeuTyrrLeuCyThrPheSerT 2456  
 DB 893 CTTTGCTGACACATGATCTGACATCTGTGCACTTCTCTGACCTGTGCACTTTTCT 952  
 QY 2456 rPalaleuLeuGluIuAlaLeuHlaSerTyrrArgAlaLeuThrGluValArgAspValAsnT 2476  
 DB 953 GGGCTCTGCTGAGGCTTGTGACCTGTACCGGGACCTCACTAGGCTGGCATGTCAACA 1012  
 QY 2476 hArgLysProMetArgPheTyrrTrpMetLeuGlyTyrrGlyValProAlaPheIleThrgLYL 2496  
 DB 1013 CCGGCCCATGCGCTTCTACTACATGCTGGGCTGGGCGTGGCTCTTCATCAACAGGGCC 1072  
 QY 2496 euAlaValAlaGluAspProGluGlyTyrrGlyAsnProAspPheCyThrLeuSerIleT 2516  
 DB 1073 TAGCGTGAGCTGTGACCCCGAGGCTAGGGAACCTTGACTTGTGTGCTGCTTCATCT 1132  
 QY 2516 YrAspThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValP 2536  
 DB 1133 ATGACAGCTCATCTGAGAGTTTGTGTGCGCGGTGGCTTTGGCTTCGATGAGTGTCT 1192  
 QY 2536 hLeuTyrrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGluYSL 2556  
 DB 1193 TCCTGTACATCTGTGCGCGCGGCTCTGTGTGCTGCCACAGCGGACGGCTTTGAGAAGA 1252  
 QY 2556 YrGlyProValSerGlyLeuGlnProSerPheAlaValIleuLeuLeuLeuSerAlaThrT 2576  
 DB 1253 AAGGCTGTGTCTGGGCGCTGAGCGCTCTTGTGGCGGTCTCTGTGTGAGCGCACGT 1312  
 QY 2576 rPLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHlaIleTyrrLeuPheAlaT 2596  
 DB 1313 GCGCTGTGCACTGTCTCTGTCAACAGCGACACCTCTCTTCCACATCACTTTGTGTA 1372  
 QY 2596 hTyraSerCyellGlnGlyProPheIlePheLeuSerTyrrValValLeuSerLysGluV 2616  
 DB 1373 CCTGAAATGTGATCAAGGCGCCCTTCATCTTCTCTATGTGTGTGTGCAAGGAGG 1432  
 QY 2616 AlArgLysAlaLeuLysLeuAlaCySerArgLysProSerProAspProAlaLeuThrT 2636  
 DB 1433 TCCGAAAGCACTCAAGCTTGTGAGCGGACCAAGCCGACCTGTCTGTGACCA 1492  
 QY 2636 hTrpSerThrLeuThrSerSerTyrrAsnCySProSerProTyrrAlaAspGlyArgLeuT 2656

```

Db      1493 CCAAGTCCACCCGACCTCGTCTACCACTGCCAGCCCTACGACGATGGCGCTGT 1552
Qy      2656 YGCIInProTyrGlyAAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerG 2676
Db      1553 ACCAGCCCTTACGAGAGACTCGGCGGCTCTGTGACAGCAACATCGCTCGGCAAGAGTTC 1612
Qy      2676 InProSerTyrIleProPheLeuLeuArgGlyLysSerAlaLeuAAspProGlyGlnGlyP 2696
Db      1613 AGCCCAAGCTACATCCCTTCTTGTCTGAGGAGAGTCCGCACTGAAACCTCGGCAAGGCG 1672
Qy      2696 roProGlyLeuGlyAAspProGlySerLeuPheLeuGlyGlnAAspGlnHisAAsp 2716
Db      1673 CCCCTGCGCTGGGCGATCCAGAGCAGCTGTCTCTGGAAGGTCAAGACAGCAACATGATC 1732
Qy      2716 roAAspThrAAspSerAAspSerAAspLeuSerLeuGlnAAspAAspGlnSerGlyTyrAlaS 2736
Db      1733 CTACACAGGAGCTCCGAGAGTGAACCTGTCTTAAAGACAGACAGAGTGTCTCTATGCTT 1792
Qy      2736 erThrHisSerSerAAspSerGlyGlyGlnGlyLysGlnGlyLysGlnGlyLysAlaAAsp 2756
Db      1793 CTACCCACCTCATCAGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1852
Qy      2756 roGlyGlnGlnGlyTyrAAspSerLeuLeuGlyProGlyAlaGlnAAspGlnProLeuHis 2776
Db      1853 CTGGAGAGCAGGCGCTGGATAGCTGCTGGGCGCTGGAGCAGAGAGAGAGAGAGAGAGAGAG 1912
Qy      2776 erThrProLysAAspGlyGlyProGlyProGlyLysAlaProTyrProGlyAAspPheGlyT 2796
Db      1913 GTATCTCCCAAGAGATGGGCGCCAGAGGCTGTGCAAGGCGCTGTGCAAGAGAGAGAGAGAG 1972
Qy      2796 hrThrAlaLysGlySerSerGlyAAsnGlyAlaProGlyLysGlnAAspGlnAAsnGlyA 2816
Db      1973 CCAAGCAAAAGAGAGTGAAGTGGCAAGGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2032
Qy      2816 spAlaLeuSerAAspGlyGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisL 2836
Db      2033 ATCCCTGTCTCAGAGAGGCTCTCCAGGCGCTCTCCAGGCTCTCTCCAGGCTCTCAACA 2092
Qy      2836 YGCIInProTyrGlyAAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerG 2856
Db      2093 AAGGCACTCTTAAAGAGAGATGTCTGCCACCATCAGCAGAGAGAGAGAGAGAGAGAGAG 2152
Qy      2856 euProLeuGlnGlyGlyThrGlySerSerArgGlySerSerAlaSerGlyLysSerArg 2876
Db      2153 TCCCGCTGAGAGCAATGCAAGGCTCTCCGAGGCTCTCCGCTAAGAGAGAGAGAGAGAG 2212
Qy      2876 YGCIInProTyrProTyrProTyrProTyrProTyrProTyrProTyrProTyrProTyr 2896
Db      2213 GCGGCGCCCTCCCTCCCGCCAGCGCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2272
Qy      2896 ecProIleAlaMetSerIleValAlaGlyThrValAAspGlnAAspSerSerGlySerGly 2916
Db      2273 TGCCCATCGCAGATGATCATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2332
Qy      2916 heLeuPhePheAAsp 2921
Db      2333 TTCTCTCTTAACTTC 2349

```

```

; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatencIn Ver. 3.1
; SEQ ID NO 636
; LENGTH: 2695
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65)..(65)
; OTHER INFORMATION: n equals a,t,g, or c
;
; NAME/KEY: misc feature
; LOCATION: (356)..(356)
; OTHER INFORMATION: n equals a,t,g, or c
;
; NAME/KEY: misc feature
; LOCATION: (2685)..(2685)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-10-264-237-636

Alignment Scores:
Pred. No.: 1,456-266 Length: 2695
Score: 281.00 Matches: 639
Percent Similarity: 97.86% Conservative: 0
Best Local Similarity: 97.86% Mismatches: 7
Query Match: 9.61% Indels: 14
DB: 12 Gaps: 0

US-09-916-849a-3 (1-2923) x US-10-264-237-636 (1-2695)

Qy      2278 ILeaThrProValValSerIleSerValHisAAspGlnGlyLeuLeuProArgAla 2297
Db      485 ATCAACACACCCGCTGTGATGATCAGCGTCCATGATGATGATGATGATGATGATGATGAT 544
Qy      2298 ILeuAAspProValThrValGlnPheArgLeuLeuGlnGlnGlnGlnGlnGlnGlnGln 2317
Db      545 CTGACAAACCCGTCAGTGAAGTTCGCTGTGAGAGACAGAGAGAGAGAGAGAGAGAGAGAG 604
Qy      2318 ILeCysValPheThrAAsnHisSerIleLeuValSerGlyThrGlyTyrTrpSerAlaArg 2337
Db      605 ATCTGTCTTCTGAGACATTCATCTGTGATGATGATGATGATGATGATGATGATGATGATG 664
Qy      2338 GlyCysGlnValValPheArgAAsnHisSerValSerCysGlnCysAAsnHisMetTh 2357
Db      665 GCGTGTGAAGTGTCTTCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 723
Qy      2357 rSerPheAlaValLeuMetAAspValSerArgArgGlnAAsnGlyLysLeuProLeuLys 2377
Db      724 GAGCTTGTGTCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 783
Qy      2377 sThrLeuThrTyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheP 2397
Db      784 GACATGACATACGAGGCTGTAGGTGCGCCTTGCGTCCCTTCTGCTCACTTCTTCTT 842
Qy      2397 heLeuThrLeuLeuArgIleLeuArgSerAAsnGlnHisGlyTyrLeuArgAAsnLeuThr 2417
Db      843 TCTCATCTCTTGTGTATCTTCGCTCCACCAACAGCGGCTCCGACCTTAACCTGACAG 902
Qy      2417 IAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyTyrLeuAlaAAspLeuProP 2437
Db      903 CTGCGCTGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 962
Qy      2437 heAla-CysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrp 2456
Db      963 TTGS-CTGACAGTATTCATTCCTGCTGACCTTCTGTACTCTGCACTTTCTCTGG 1021
Qy      2457 AlaLeuLeuGlnAlaLeuHisLeuTyrArgAlaLeuThrGlnValArgAAspValAAsnThr 2476
Db      1022 GCTCTGTGAGAGGCTGTGACCTGTACCGGAGCACTCAAGAGTGGCGATGTCAACACC 1081
Qy      2477 GlyPromerArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeu 2496
Db      1082 GGCCCATCGCTTCTACTACATGCTGGGCTGGGCGTGGCTGCTTCTATCAACAGGCGTA 1141

```

```

RESULT 11
US-10-264-237-636
; Sequence 636, Application US/10264237
; Publication No. US2004009491A1
; GENERAL INFORMATION:
; APPLICANT: Biree et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19

```





[illegible]

Db	1615	GCTGTAACAGCCCTTACCGAAGACTGGCCGGCTCTCTGCAACAGCACAGTGGCTGGGGCA	1674
Oy	2674	SSerGlnProSerTyrIleProPheLeuArgGluGluSerAlaLeuAnProGly	2694
Db	1675	GAGTCACGCCAGCTACATCCCTTCTTGAGGGAGGAAGATCCGCACTGAACCTGGCCCA	1734
Oy	2694	nglyProProGlyLeuGlyIAspProGlySerLeuPheLeuGluGlyValIAspGlnGlnI	2714
Db	1735	AGGGCCCCCTGGCTGGGGGAGTCCAGGAGCGCTGTCTTGAAAGGTCAAGACAGACAGCA	1794
Oy	2714	SASPProAspThrAspSerAspSerAspLeuSerLeuGluIAspAspGlnSerGlySerTyr	2734
Db	1795	TGATCTCGACACGGAGCTCCGACAGTGACCTGTCTTGAAGAAGACCAAGATGGCTCTTA	1854
Oy	2734	ralSerThrHisSerSerAspSerGluGluGluGluGluGluGluGluGluGluGluGluGlu	2754
Db	1855	TGCTCTTACCCATCCATCAAGCAGTGAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAG	1914
Oy	2754	APheProGlyValGluGlnGlyTyrAspSerLeuLeuGlyProGlyValIAspGluArgLeuProle	2774
Db	1915	CTTCCCTCGAGAGAGAGGGCTGGGATACCTGCTGGGGCTCGAGACAGAGAGACTGGCCCT	1974
Oy	2774	uHisSerThrProIAspGlyGlyProGlyProGlyValAspAlaProTyrProGlyIAspPh	2794
Db	1975	GCAAGACTCCCAAGATGGGGGCCCAAGGGCCCTGGCAAGGCCCTGGGCAAGAGACTT	2034
Oy	2794	eglyThrThrAlaArgGluSerSerGlyIAspGlyAlaProGluGluIAspIAspArgGluAs	2814
Db	2035	TGGAGCAACAGCAAAAGAGAGTATGGCAACGGGGCCCTGAGAGAGCGGCTGGGGAGAA	2094
Oy	2814	nglyAspAlaLeuSerArgGlySerLeuGlyProLeuProGlyIAspSerAlaGlnPr	2834
Db	2095	TGAGAGATGCCCTGCTCGAGAGGGGTCCCTAGGGCCCTTCAAGGCTCTTCTGGCCAGCC	2154
Oy	2834	ohIAspGlyIleLeuIAspIAspGlySerGluProThrIleSerGluIAspSerLeuLe	2854
Db	2155	TCACAAAGAGCATCTTAAAGAAAGTGTGCCCCACATCAGCAGAGAAAGACAGCTCTCT	2214
Oy	2854	uArgLeuProLeuGluGlnGlyCysThrGlySerSerArgGlyIAspSerAlaSerGluGlySe	2874
Db	2215	GCGGCTCCCTGGAGCAATGACAGAGGTCTTCCGGGGCTCCTCGCTAGTGAAGGAG	2274
Oy	2874	raArgGlyGlyProProProArgProProProArgGlnSerLeuGlnGluIAspGlnAsnG	2894
Db	2275	CCGGGGGGGK-CCCCCTCCCCCGCCACCGCCCGGAGAGACCTCCAGAGACAGCTGAAG	2333
Oy	2894	IValMetProIleAlaMetSerIleValIAspGlyThrValIAspGluAspSerSerGlyS	2914
Db	2334	GGGTCATGCCCATGCGCATAGCATCAAGCAGGCAAGCGATGATAGGACTTCGTCAAGCT	2393
Oy	2914	erGluPheLeuPheAsnPheAsnIAspLeuHis	2923
Db	2394	CCGAATTCCTCTTAACTCCTGAT	2422
RESULT 13			
US-09-764-870-571			
; Sequence 571, Application US/09764870			
; Patent No. US20020042386A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: PT214			
; CURRENT APPLICATION NUMBER: US/09764,870			
; PRIORITY FILING DATE: 2001-01-17			
; Prior application data removed - consult PALM or file wrapper			
; NUMBER OF SEQ ID NOS: 646			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 571			
; LENGTH: 652			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-764-870-571			

## Alignment Scores:

Pred. No.: 5,93e-204 Length: 652  
 Score: 217.00 Matches: 217  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.42% Indels: 0  
 DB: 9 Gaps: 0

US-09-916-849A-3 (1-2923) x US-09-764-870-571 (1-652)

```

QY 567 GluLeuAspArgGluGluValAspPheTyrSerPheGlyValAlaArgAspHisGly 586
DB 1 GAACGGACCGGAGAGAAATTGATTCTACAGCTTGCGGTAGAAAGCTGAGACCACTGAC 60
QY 587 ThrProAlaLeuThrAlaSerAlaSerValThrValLeuAspValAsnAspAsn 606
DB 61 ACTCCAGCACTACAGCTCGCGCAGTGTACAGGTGACTGTCTCTGATGTCAACAGCAAC 120
QY 607 AsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaAlaValGly 626
DB 121 AATCCAACTTTAACCAACAGATACAGTCCGCTCAATGAGATGCACTGTGGGC 180
QY 627 ThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleThrTyrGln 646
DB 181 ACCAGCGGTGACGGGTGTCACTGTGACCGGTGATGCTCATAGTGTCACTACCTACAG 240
QY 647 IleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGlyLeu 666
DB 241 ATCCACAGTGGCAATACCTGAAACCGCTTCTCATACCAAGCAAGGTGTGGTGGCTG 300
QY 667 ValSerLeuAlaLeuProLeuAspTyrIleuGluArgGlnTyrValLeuAlaValThr 686
DB 301 GTATCCCTTGCCCTGCGCACTGAGCTACAACTTGAGCGGCAATGTGTGGCTGTACC 360
QY 687 AlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAspAlaAsn 706
DB 361 GCTCCGAGTGGCACTCGGACAGACCGGCAAGATTGTGTGAATGTCAACCAACCGCAAC 420
QY 707 ThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAspArgPro 726
DB 421 ACCCATCGCTGCTGCTTTCAGAGCTCCCACTACAGTAAAGTAAATGAGAGACCGGCG 480
QY 727 AlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArg 746
DB 481 GCAAGCAACACGGTGGTGTGATCAGCGCCACGATGAGGACACAGGTGAGAAATCCCGC 540
QY 747 IleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAla 766
DB 541 ATCACTTACTTTCATGAGAGACAGATCCCGCAGTTCCGATCGATGACAGACCGGGGCT 600
QY 767 ValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeu 783
DB 601 GTCAACCAACCGGCTGAGCTGAGCTATGAAACCAAGTGTCTTACACCTG 651

```

## RESULT 14

US-10-125-540-571

Sequence 571, Application US/10125540  
 Publication No. US20030059875A1  
 GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT214C1  
 CURRENT APPLICATION NUMBER: US/10/125,540  
 CURRENT FILING DATE: 2002-04-19  
 Prior Application removed - See File Wrapper or Palm  
 NUMBER OF SEQ ID NOS: 646  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 571  
 LENGTH: 652  
 TYPE: DNA

ORGANISM: Homo sapiens  
 US-10-125-540-571

## Alignment Scores:

Pred. No.: 5,93e-204 Length: 652  
 Score: 217.00 Matches: 217  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.42% Indels: 0  
 DB: 15 Gaps: 0

US-09-916-849A-3 (1-2923) x US-10-125-540-571 (1-652)

```

QY 567 GluLeuAspArgGluGluValAspPheTyrSerPheGlyValAlaArgAspHisGly 586
DB 1 GAACGGACCGGAGAGAAATTGATTCTACAGCTTGCGGTAGAAAGCTGAGACCACTGAC 60
QY 587 ThrProAlaLeuThrAlaSerAlaSerValThrValLeuAspValAsnAspAsn 606
DB 61 ACTCCAGCACTACAGCTCGCGCAGTGTACAGGTGACTGTCTCTGATGTCAACAGCAAC 120
QY 607 AsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaAlaValGly 626
DB 121 AATCCAACTTTAACCAACAGATACAGTCCGCTCAATGAGATGCACTGTGGGC 180
QY 627 ThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleThrTyrGln 646
DB 181 ACCAGCGGTGACGGGTGTCACTGTGACCGGTGATGCTCATAGTGTCACTACCTACAG 240
QY 647 IleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGlyLeu 666
DB 241 ATCCACAGTGGCAATACCTGAAACCGCTTCTCATACCAAGCAAGGTGTGGTGGCTG 300
QY 667 ValSerLeuAlaLeuProLeuAspTyrIleuGluArgGlnTyrValLeuAlaValThr 686
DB 301 GTATCCCTTGCCCTGCGCACTGAGCTACAACTTGAGCGGCAATGTGTGGCTGTACC 360
QY 687 AlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAspAlaAsn 706
DB 361 GCTCCGAGTGGCACTCGGACAGACCGGCAAGATTGTGTGAATGTCAACCAACCGCAAC 420
QY 707 ThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAspArgPro 726
DB 421 ACCCATCGCTGCTGCTTTCAGAGCTCCCACTACAGTAAAGTAAATGAGAGACCGGCG 480
QY 727 AlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArg 746
DB 481 GCAAGCAACACGGTGGTGTGATCAGCGCCACGATGAGGACACAGGTGAGAAATCCCGC 540
QY 747 IleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAla 766
DB 541 ATCACTTACTTTCATGAGAGACAGATCCCGCAGTTCCGATCGATGACAGACCGGGGCT 600
QY 767 ValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeu 783
DB 601 GTCAACCAACCGGCTGAGCTGAGCTATGAAACCAAGTGTCTTACACCTG 651

```

## RESULT 15

US-09-764-870-568

Sequence 568, Application US/09764870  
 Patent No. US20020042386A1  
 GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT214  
 CURRENT APPLICATION NUMBER: US/09/764,870  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 646  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 568  
 LENGTH: 509  
 TYPE: DNA

ORGANISM: Homo sapiens  
 US-09-764-870-568

## Alignment Scores:

Pred. No.:	3,87e-119	Length:	509
Score:	131.00	Matches:	168
Percent Similarity:	99.41%	Conservative:	0
Best Local Similarity:	99.41%	Mismatches:	1
Query Match:	4.48%	Indels:	1
DB:	9	Gaps:	0

US-09-916-849a-3 (1-2923) X US-09-764-870-568 (1-509)

```

QY      389  ValGluAspAspAsnAspAsnAlaProGlnPheSerGluLysArgTyrValValGlnVal 408
      |||
Db      2    GTGAGAGATGACATATATATGCCCCCAGTTTATGAGAGAGCGCTATGTGTCAGGTG 61

QY      409  ArgGluAspValThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspLys 428
      |||
Db      62  AGGAGAGATGTGCTCCAGGGGCCCGAGTCTCCAGTCCAGGCTTGAGTCAGAGAGAG 121

QY      429  GlySerAsnAlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyr 448
      |||
Db      122  GGGAGCAATGCCGTGGTGCATATAGCATCATGATGCGCAATCTCGGGGACAGTTTAT 181

QY      449  LeuAspAlaGlnThrGlyAlaLeuAspValValSerProLeuAspTyrGluThrThrLys 468
      |||
Db      182  CTGGATGCCCACTGGAGCTCTGATGTGTGAGGCCCTCTTGACTATGAGACGACCAAG 241

QY      469  GluTyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSer 488
      |||
Db      242  GAGTACACCTTACGGGTGCGAGACACAGATGTGGCCGCTCCCACTCTTATATGTCT 301

QY      489  GlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThr 508
      |||
Db      302  GGCTGTGTACAGTACAGTCTGTGATATCAACGACATGCCCATCTTCGTCAAGACC 361

QY      509  ProPheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGln 528
      |||
Db      362  CTTTCCAGGCTACTGCTGAGAGATGT-CCCTTAGGCTAAGTCTGTTCCATGTCCAG 420

QY      529  AlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGly 548
      |||
Db      421  GCTATCGACGCTGATGTGTGACAAATGCCCGCTGGAATACCGCTTGTCTGGGGTGGGA 480

QY      549  HisAspPheProPheThrIleAsnAsn 557
      |||
Db      481  CATGACTTCCCTTCAACATCAAT 507

```

Search completed: February 15, 2004, 03:04:58  
 Job time : 2287 sec8

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 10:45:04 ; Search time 11526 Seconds  
(without alignments)  
6163.629 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 2923  
Sequence: 1 MNSPATGVFLPFPPLLL.....ACTVDEDSGSEFLFPFLH 2923

Scoring table: Oligo  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DBV=xlh  
-O=/cgn2\_1/USPTO.spool/US09916849/runat\_11022004\_143917\_21362/app\_query.fasta\_1.3079  
-DB=EST -QFMT=fastap -SUFFIX=oligo -MIMATCH=0.1 -LDOPCD=0 -LDOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849@cgn2\_1 1 9354 @runat\_11022004\_143917\_21362 -NCPU=6 -ICPU=3  
-NO.WMAP -LARGEBUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGIDC  
-BEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELCP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	9.4	907	13 BX409742	BX409742 BX409742
2	252	8.6	896	13 BU173510	BU173510 AGENCOURT
3	235	8.0	1076	12 BM480097	BM480097 AGENCOURT
4	226	7.7	908	13 BU857213	BU857213 AGENCOURT
5	210	7.2	695	9 AU116945	AU116945 AGENCOURT
6	204	7.1	662	9 AU042580	AU042580 DKF2P434K
7	204	7.0	877	13 BX409741	BX409741 BX409741
8	199	6.8	906	13 BQ677400	BQ677400 AGENCOURT
9	191	6.5	927	13 BQ679970	BQ679970 AGENCOURT
10	188	6.4	930	13 BQ676596	BQ676596 AGENCOURT
11	187	6.4	893	13 BU849091	BU849091 AGENCOURT
12	179	6.1	640	9 AW370778	AW370778 QV0-BT026
13	169	5.8	556	10 BE008566	BE008566 RCS-BN015
14	166	5.7	523	13 BX110290	BX110290 BX110290
15	165	5.6	549	10 BP951925	BP951925 QV1-NN022
16	158	5.4	477	9 AL120631	AL120631 DKF2P751I
17	151	5.2	465	10 BE012097	BE012097 RCS-BN105
18	151	5.2	498	9 AL704729	AL704729 DKF2P686H
19	151	5.2	816	13 BU701387	BU701387 UI-M-EXO-
20	148	5.1	647	13 BU703750	BU703750 UI-M-FOO-
21	142	4.9	990	13 BX460356	BX460356 BX460356
22	141	4.8	483	10 BE009645	BE009645 PMA-BN017
23	139	4.8	446	12 BI032886	BI032886 MR4-NN020
24	127	4.3	597	14 CA891550	CA891550 B0166E02
25	127	4.3	1018	13 BQ430341	BQ430341 AGENCOURT
26	124	4.2	405	10 BP958428	BP958428 RC4-NN117
27	122	4.2	425	10 BF092095	BF092095 RCO-TN007
28	121	4.1	616	9 AW748496	AW748496 QV0-BT026
29	115	3.9	500	2 HSM085357	HSM085357 Homo sapi
30	115	3.9	627	12 BI456703	BI456703 Homo sapi
31	115	3.9	722	12 BI456543	BI456543 603172181
32	114	3.9	738	14 CA510833	CA510833 UI-R-PJ0-
33	112	3.8	812	12 BG915872	BG915872 602815323
34	111	3.8	334	14 CB122348	CB122348 K-BST0170
35	109	3.7	480	2 HSM073596	HSM073596 Homo sapi
36	107	3.7	557	10 BF369920	BF369920 UI-R-B52-
37	106	3.6	455	10 AW900873	AW900873 CML-NN100
38	104	3.6	918	14 CA454005	CA454005 AGENCOURT
39	102	3.5	598	9 AL704278	AL704278 DKF2P686O
40	100	3.4	310	10 BE703584	BE703584 MR2-NN111
41	100	3.4	506	14 CB783412	CB783412 AMGNIC.N
42	100	3.4	694	14 CD350873	CD350873 UI-M-GTO-
43	100	3.4	846	14 CA751074	CA751074 UI-M-FOO-
44	98	3.4	621	10 AW955385	AW955385 EST367455
45	97	3.3	466	10 BF915504	BF915504 MR3-UT012

## ALIGNMENTS

RESULT 1  
LOCUS BX409742 907 bp mRNA EST 13-MAY-2003  
DEFINITION BX409742 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
ACCESSION CSDDP013YE18 5-PRIME, mRNA sequence.  
VERSION BX409742.1 GI:30653001  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Butheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 907)

AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 131 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 64.f. For more  
 information about this cluster see http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BAF0122G07\_AF01112\_2&cluster=64.f.  
 Contact: Feng Liang Email: fliang@life.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0BAF0122G07\_AF01112\_2.  
 Location/Qualifiers

## FEATURES

source  
 1..907  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF013YE18"  
 /library="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-Oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 BASE COUNT 148 a 322 c 238 g 198 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 5,49e-236 Length: 907  
 Score: 275.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 9.41% Indels: 0  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BK409742 (1-907)

QY 2433 AAlaApleuProPhelaACGTTThValIleAlaIleleuLeuHsPhleuTyTleuCyB 2452  
 Db 7 GCTGACCTCCCTTGGCTGCAAGTCATTCCTGCTGCACTTCTGACCTGCTGCTG 66  
 QY 2453 ThPheserTpaAlaLeuGluAlaLeuHsleuTyTpaAlaLeuThrgiValArg 2472  
 Db 67 ACCCTTCTCGGGCTCTGCTGAGAGCTTGGACCTGACCGGGCACTCACTGAGGCGC 126  
 QY 2473 AspValaAnthrgiYProMetArgPheTyTymeLeuGlyTTrgiValProAlaPhe 2492  
 Db 127 GATGTCAACACCGGCCCCCATGCGCTTCACTACATGCTGGGCGGCGCTGCTTC 186  
 QY 2493 IleThrgiYleuAlaValGlyLeuAspProGluGlyTTrgiValPProAlaPheCgTTP 2512  
 Db 187 ATCAACAGGGCTAGCGGCTGAGACCCGAGGGCTACAGGAACCTGCTGCTG 246  
 QY 2513 LeuSerIleTyTpaThleuIleTTPSerPheAlaGlyProValAlaPheAlaValSer 2532  
 Db 247 CTCTCCATCTATGACAGCTCATCTGAGATTGCTGCGCCGGTGGCTTTCGCTCG 306  
 QY 2533 MetSerValPheLeuTyTleuAlaAlaArgAlaSerCyValAlaGlnATrgingly 2552  
 Db 307 ATGAGTCTTCTTCGTCATCTGCGCGGCGGCTCTGCTGCTGCCAGCGAGG 366  
 QY 2553 PheGluTyTpaGlyProValSerGlyLeuGlnProSerPheAlaValleuLeuLeu 2572  
 Db 367 TTGTAGAGAAAGGTCTGCTGCTGCGGCTGCGACCTCTTCGCTGCTGCTG 426  
 QY 2573 SerAlaThrTTrpLeuAlaAlaLeuSerValaMserAphTTrpLeuAlaPheIleTyT 2592  
 Db 427 AGGCGCACGTGGCTGCTGCGACGCTCTCTGTCAACAGGACACCTCTCTTCACCTAC 486

QY 2593 LeuPheAlaThrCyAsnCysIleGlnGlyProPheIlePheLeuSerTyTValValleu 2612  
 Db 487 CTCTTGTACTGACTGCAATTCATCCAGGGCCCTTCATCTCTCTCATATGATGCT 546  
 QY 2613 SerTyGluValArgValAlaLeuLysLeuAlaCysSerArgLyProSerProAlaPro 2632  
 Db 547 ACCAAGAGGTCCGGAAGACCTCAAGCTTGTCTGACCGGACCCGACCTCAACCT 606  
 QY 2633 AlaLeuThrThySerThreThrsSerSerTyTpaCysPProTyTpaAlaP 2652  
 Db 607 GCTTGACCAACCAAGTCCACCTGACCTGCTCTCAACATGCCCCAGCCCTACGAGAT 666  
 QY 2653 GlyArgLeuTyTglnProTyTglnAspSerAlaGlySerLeuHsSerThSerArgSer 2672  
 Db 667 GGGCGGCTGTACCAACCTACAGGACATCGCGGCTCTCTGACAGACAGACAGTGGCTCG 726  
 QY 2673 GlyLysSerGlnProSerTyTleProPheLeuAlaGluSerAlaLeuAlaP 2692  
 Db 727 GGCAGAGTCAAGCCACAGCTACATCCCTCTTCTGAGAGGAGAGATCCGCACTGAACCT 786  
 QY 2693 GlyGlnGlyProGlyLeuGlyAlaPProGlySerLeuPheLeu 2707  
 Db 787 GGCAGAGGCCCCCTGCGCTGGGGATCCAGGACCTGTCTCTG 831

RESULT 2  
 BU173510 896 bp mRNA linear EST 04-SRP-2002  
 LOCUS BU173510  
 DEFINITION AGENCOURT\_7939348 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6172948  
 5', mRNA sequence.  
 ACCESSION BU173510 GI:22687494  
 VERSION BU173510.1 GI:22687494  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapds-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM13543 row: 1 column: 05  
 High quality sequence stop: 677.

## FEATURES

source  
 1..896  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6172948"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_67"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Library constructed by Life  
 Technologies."

BASE COUNT 169 a 282 c 267 g 177 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.51e-215 Length: 896  
 Score: 252.00 Matches: 252  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.62% Indels: 0

DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BU173510 (1-896)

QY 1351 PheLYsYsAAsPcYsPProSeGlyAAsPheGlyLysProLYrCYsGluValThrThrArg 1370

Db 1 TTCAGATGCGATGGCCATCTGGAGACTTCGAGAGCCCTACGCAAGGTGACACGCGC 60

QY 1371 SerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThr 1390

Db 61 AGCTTCCCCGCCCACTCTTCATCACTTGGCGGCGCGCAAGCTTCCACTTCACTTCAAC 120

QY 1331 LeuAlaLeuSerPheAlaThrIleGlyLysGlyLeuLeuLeuLYrAAsGlyAArgPhe 1410

Db 121 CTGGCCCTCTCGCTTGGCCCAAGAGAGCGCGCGCTGCTGTGTACATGGCGGTTTC 180

QY 1411 AAsGlyLysHisAAsPheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPhe 1430

Db 181 AATGAGAAAGCATGACTTTGTGGCCCTCGAGGTGATCCAGAGCAGGCTCAAGCTTTC 240

QY 1431 SerAlaGlyLysSerThrThrThrValSerProPheValProGlyGlyValSerAAsGly 1450

Db 241 TCTGCAAGGGAGTCAACACCAACGAGTGTCCCATTCGTGCCGAGAGTCAGTATGGC 300

QY 1451 GlnTPHISThrValGlnLeuLysLYrTYrAAsLYrProLeuLeuGlyGlnThrGlyLeu 1470

Db 301 CAGTGGCATACGCTGACGTGCAATATCACTAATAGCACTGTGGTCAAGCAGGCGTTC 360

QY 1471 ProGlnLYrProSeGlyLysGlyValAlaValIleThrValAAsPGLYrCYsAAsPThrGly 1490

Db 361 CCACAGGCGCCATTCAGAGCAGAGAGTGGCTGTGGAGACGAGTGGAGCTGTGACACAGAA 420

QY 1491 ValAlaLeuArgPheGlySerValLeuGlyAAsLYrSerCYsAlaAlaGlnGlyThrGln 1510

Db 421 GTGGCCTTGGCGTGGATCTGTCTCGGGCAACTCTCTGTGCTGCCAGGGCAACCGAG 480

QY 1511 GlyGlySerLYrLysSerLeuAAsPLeuThrGlyProLeuLeuGlyGlyValProAAsP 1530

Db 481 GGTGGACACCAAGAAATCTCTGATCTGACGGGCGCCCTGCTACTAGCGGGGTGTCTTAC 540

QY 1531 LeuProGlyLysSerPheProValArgPheArgGlnPheValGlyCYsMetAAsAAsLeuGln 1550

Db 541 CTGCCCGAGAGCTTCCGAGTCCGAAATGGCGGCACTGTGGCTGCAATGGCACTGCAAG 600

QY 1551 ValAAsSerArgHisIleAAsPheAlaAAsPheIleAlaAAsAAsGlyThrValProGly 1570

Db 601 GTGACACACCGGCACATAGACATGGCTGACTTCACTTGCACAAAGGCAACGCTGCGC 660

QY 1571 CysPProAlaLYrLysAAsValCYsAAsSerAAsThrCYsHisAAsGlyGlyThrCYsVal 1590

Db 661 TGCCCTGGCCAGAAAGAACGTGTGTGACAGCAACCTTGCACAAATGGGGGCACTTGGCGTG 720

QY 1591 AAsGlnTPRAsPAlaPheSerCYsGlyCysProLeu 1602

Db 721 AACCACTGGAGCGCTTCACTGCTGCAAGTCCCCCTG 756

RESULT 3

LOCUS BM480097 1076 bp mRNA linear EST 05-FEB-2002

DEFINITION AGENCOURT\_6468579 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5574047

ACCESSION BM480097

VERSION BM480097.1 GI:18529139

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1076)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>

Plate: L1AM12322 row: a column: 24

High quality sequence start: 29

High quality sequence stop: 666.

FEATURES

source

1..1076

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5574047"

/ribose\_type="duodenal adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH MGC 88"

/note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."

BASE COUNT 215 a 357 c 310 g 193 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 5.53e-200 Length: 1076

Score: 235.00 Matches: 275

Percent Similarity: 99.28% Conservative: 0

Best Local Similarity: 99.28% Mismatches: 1

Query Match: 8.04% Indels: 2

DB: 12 Gaps: 0

US-09-916-849a-3 (1-2923) x BM480097 (1-1076)

QY 1942 GlnCYsProCYsLYrProGlyValIleGlyAArgGlnCYsAAsPArgCYsAAsPProPhe 1961

Db 38 CAGGTGTCATGACCAAGAGGTGATCGGCGTCAAGTGTACACGCTGTGACAACTTTT 97

QY 1962 AlaGluValThrThrAAsGlyCYsGluValAAsTYrAAsPserCYsProArgAlaIleGlu 1981

Db 98 GCTAGGTGACCAACCAATGGCTGTGAAGTATATACAGCTGCCACGACGACATTGAG 157

QY 1982 AlaGlyIleTPRTPRProArgTYrArgPheGlyLeuProAlaAlaAAsProLYr 2001

Db 158 GCTGGAAATCGTGGGCCCCGTAACCGCTTGGGCTGCTGTGCTCTCCCTGCCAAA 217

QY 2002 GlySerPheGlyThrAlaValArgHisCYsAAsPGLYrHisAArgGlyTYrLeuProProAAsn 2021

Db 218 GGCCTCTTGGAGCTGCTGTGCCCACTGTGATAGACACAGGGGTGGCTCCCCCAAC 277

QY 2022 LeuPheAAsCYrThrSerIleThrPheSerGlnLeuLysGlyPheAlaGluArgLeuGln 2041

Db 278 CTCTTCAACTGACAGCTCACTCACTTCTCAGAACTGAAGGGCTTCGCTGAGCGGCTACAG 337

QY 2042 ArgAAsGlySerGlyLeuAAsPserLYrArgSerGlnLeuAlaLeuLeuAAsPAsn 2061

Db 338 CGGAATAGTCAAGGCTTCAAGCTCAGGGCGCTCCAGAGCTTGAAGCTTCTGCGCAAC 397

QY 2062 AlaThrGlnHisThrAlaGlyTYrPheGlySerAAsPValLYrValAlaTYrGlnLeuAla 2081

Db 398 GCGAAGGAGCAACAGCTGGCTACTTGGGAGCGACACTCAAGGTGGCTCAAGCTGGCC 457

QY 2082 ThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAAsP 2101

Db 458 ACGCGGCTGTGGGCCACAGAGACACCAAGCGGGCTTGTGGCTGTGCGCACACAGAC 517

QY 2102 ValHisPheThrGluAAsLeuLeuArgValGlySerAlaLeuLeuAAsPThrAlaAAsLYr 2121

Db 518 GTGCACCTTCACTAGAAATCTGTGGGGGTGGGAGCGCCCTCTGTGACACAGCCAAAG 577



QY	2122	ArgHisrPbGluLeuIleGIgInthrGluGluYgYlrrAlrPLeuLeuGIrHrIstYr	2144
Db	578	CGGCACrTGGAGCTGATrTCCAGCACrAGAGGGrTGGCACCGCTGGCTCTCCAGACrATrT	637
QY	2142	GIuAlArYrAlAsErAlAlEuAlAGInArMeArGHisrYrYrLeuSerProPheThr	2161
Db	638	GAGCGCTrACGGCCAGrGCCCCrTGGCCCGACrAACTGCGCACrCTrACCTrAGCCCCrCTTACCC	697
QY	2162	ILeValThrProArMileValIleSerValValArgLeuAsp-LysGILyAsnPheAlaGI	2181
Db	698	ATrGTCCrAGCCCACTrTGTCTrCATrCTCCGTAGrGGCGCTTGGACrAANAAG-AACTTTCGTGG	756
QY	2181	YAlAlYsLeuProArGTYrGILuAlAlEuArGILyGluGInProPArPLeuGluThrTh	2201
Db	757	GGCCrAAGCTGGCCCCGCTrACGAGrGCCCCrCTGTrGGGAGrAGrGCCCCCGGACrCTTGGAGrCAAC	816
QY	2201	rValIleLeuProGluSerValAlPheArGILyThrProProValValArg	2217
Db	817	AGrCATrTTCGCTTGAGrGTCTGTCTTTCGAGrAGrAGrGCCCCCGTGTGTCAAG	865

LOCUS	908 bp	mRNA	linear	EST
LOCUS	908 bp	mRNA	linear	EST
DEFINITION	AGENCOURT 10480787 NTH MGC_107 Homo sapiens	clone		
IMAGE:6647022 5', mRNA_sequence.				

ACCESSION	B085/413	
VERSION	B0857213.1	GI:24042203
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 908)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.

Email: [cgaps-i@mail.nih.gov](mailto:cgaps-i@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MSC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:

http://image.llnl.gov  
Plate: LLCM2886 row: m column:  
High quality sequence stop: 646

FEATURES	Location/Qualifiers
source	1. .908

```

/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:664702"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

```

/clone\_id=NIH\_MGC\_107" /notes=Organ: breast; Vector: POT87; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library.

BASE COUNT	190 a	300 c	281 g	137 c
ORIGIN				

Alignment Scores:

Pred. No.:	5.83e-192
Score:	226.00

Percent Similarity: 100.00%  
Best Local Similarity: 100.00%

Query Match:	7.73%	Indels:	0
DB:	13	Gaps:	0
US-09-916-849A-3 (1-2923) x BUB57213 (1-908)			

2654 ArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGly 2673  
Db 2 CGGCTGACCAAGCCCTACGGAGACTCGGCGGCTCTCTGCACAGCACCAAGTCGCTCGGGC 61

Db 62 AGAGTAGCCACGCTACATCCCTTCTTGCTGAGGGAGGAGTCGCGACTGAACCTGGC 121

[illegible]

2 182 CATGATCTGACACGAGCTCCACAGTACCTGTCTTAGAAGACGACCAAGACTGGCTTC 241

Db 242 TATGCTTACCCACTCATCAGACGTGAGGAGGAGAGAGAGAGAGAGAGGCC 301

302 GCCTTCCCTGGAGAGAGGGGCTGGGATAGCCCTGCTGGGGCCCTGGAGCAGAGACATCGCC 361

Db 362 CCGCACAGTACTCCCAAGGATGGGGGCCCGAGGGCTGTGGCAAGGCCCTCTGGCCAGAGAC 421

Db 422 TTTGGGACCAAGCAAAGAGGTAGTGGCAACGGGGCCCTTAGAGAACGGCTGGCGGAG 481

Db 482 AATGAGAGTGGCTGTCTTCGAGAGGGGCTCCATAGGCCCCCTTCAGGGCTTCTGCCAG 541

2554 *LeuArgIeuProLeuIuIuIncyIeThcIyIseIerSerIrgIyIseIerIaIseIrgIuIyI* 2873  
 Db 542 *CCTCAAGAGGATCCTTAAGAGAGAAAGTGTCTGCCACCATCAGCGAAGAGACAGCCTC* 601  
 Qy

Db 602 CTGCGGCTCCCCCTGGAGCAATGACACAGGGTCTCCCGGGGCTCTCCGCTAGTGAGGAC 661

Gy 2874 SerrArgIyglyProPro 2879  
 ||||| ||||| ||||| |||||

DD	66Z	AGCCGGGCGGCCCCC	6/9
RESULT 5			
AV116945			

[illegible]

KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCES  
Mammalia; Eutheria; 1 (bases 1 to 695)  
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, I., Wakabayashi, T., Nakamura, Y., Nagai, T., Sugano, S. and

TITLE  
JOURNAL  
COMMENT  
ISOGAI, T.  
HRI human CDNA project  
Unpublished  
Contact: Takao Isogai

GENOMICS LABORATORY

Helix Research Institute  
1532-3 Yana, Kibarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

## FEATURES

source

1. 695  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEMBA1000310"  
/issue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/clone\_1ib="HEMBA1"  
/note="Vector: pME18SFL3"  
BASE COUNT 105 a 246 c 174 g 167 t 3 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.1e-177 Length: 695  
Score: 210.00 Matches: 210  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.18% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AU116945 (1-695)

QY 2399 ThrLeuLeuArgIleLeuArgSerAngInHsgIyIleArgArgAsnLeuThraIaIa 2418  
DB 1 ACTCTTCGCGATCTCGCGCTCAACCAACGCGCATCGACGTAACTGACAGCTGCC 60  
QY 2419 LeuGIyLeuAlaGlnLeuValPheLeuLeuGIyIleAngInAlaAspLeuProPheAla 2438  
DB 61 CTGGGCGCTGGCTCAGCTGGCTCTTCCTCGGAAATCAACAGGCTGACCTCCCTTTGGC 120  
QY 2439 CysThrValIleAlaIleLeuLeuHisPheLeuTyLeuCysThrPheSerTrpAlaLeu 2458  
DB 121 TGCAAGATATGCGCATCTGCTGCACTTCCTGACCTTTCCTGGGCTCTG 180  
QY 2459 LeuGluAlaLeuHisLeuTyArgAlaLeuThrgIuValArgAspValAsnThrgIyPro 2478  
DB 181 CTGGAGGCGCTTCACCTGACCGGGCACTCACTGAGGTGCGGATGTCAACCGGCCCC 240  
QY 2479 MetArgPheTyTrpMetLeuGIyTrpGIyValProAlaPheIleThrgIyLeuAlaVal 2498  
DB 241 ATGCCCTTCTACTCAATGCTGGGCTGGGCGGTGGCTGCTTCATCAAGGGCTAGCCGTG 300  
QY 2499 GIyLeuAspProGIyGIyTyArgIyAsnProAspPheCysTrpLeuSerIleTyArgPthr 2518  
DB 301 GGCGCTGACCCGAGGGGCTACCGGAACTTGACTTCTGCTGCTCTCCATCTATACACG 360  
QY 2519 LeuIleTrpSerPheAlaGIyProValAlaPheAlaValSerMetSerValPheLeuTy 2538  
DB 361 CTCACTCTGAGATTGTTGCTGGCGGGTGGCTTGGCGTCTCGATGAGTGTCTTCCGTAC 420  
QY 2539 IleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGIyPheGluIyLeuGIyPro 2558  
DB 421 ATCTGGCGGGCGCGGCGCTCTGCTGTGCTGCCAGCGGCAAGGCTTTGAGAAAGATCTT 480  
QY 2559 ValSerGIyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeu 2578  
DB 481 GTCCTGGGGCTGACAGCCCTCTTGGCGGTCTCTGCTGCTGAGGCCACGCGGCTGCTG 540  
QY 2579 AlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyLeuPheAlaThrCysAsn 2598  
DB 541 GCACGTCTCTCTGTCAACAGCAACCCCTCTTCCACTACCTCTTGTCTACCTGCAAT 600  
QY 2599 CysIleGlnGIyProPheIlePheLeuSer 2608

DB 601 TGCAATCAGGGCGCTTCATCTCTCTCC 630

## RESULT 6

AL042580/c 662 bp mRNA linear EST 29-FEB-2000  
LOCUS DKFZp434K0121\_r1 434 (synonym: htes3) Homo sapiens cDNA clone  
DEFINITION DKFZp434K0121 5', mRNA sequence.

ACCESSION AL042580  
VERSION AL042580.1 GI:5422032  
KEYWORDS EST  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 662)  
AUTHORS Blum H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
TITLE EST (Blum, et al.)  
JOURNAL Unpublished  
COMMENT Unpublished  
CONTACT: Blum H  
MIPS

Am Klopferplatz 18a D-8152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by LMU (Ludwig Maximilians University,  
Munich/Germany) within the cDNA sequencing consortium of the German  
Genome Project.

No sl sequence available.  
This clone (DKFZp434K0121) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source

1. 662  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp434K0121"  
/issue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="MDH10B"  
/clone\_1ib="434 (synonym: htes3)"  
/note="Vector: pSPORT1, Site\_1: NotI, Site\_2: SalI"  
BASE COUNT 148 a 197 c 179 g 138 t  
ORIGIN

## Alignment Scores:

Pred. No.: 5.24e-175 Length: 662  
Score: 207.00 Matches: 207  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.08% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AL042580 (1-662)

QY 380 SerThrThraIaAlaValPheLeuSerValGIuAspAsnAspAlaProGlnPhe 399  
DB 623 AGTACCAAGCGCGTGTTCCTTCTGTGAGAGATGACATGATATGCCCCCAAGTTT 564  
QY 400 SerGluYsaArgTyValValGlnValArgGluAspValThrProGIyAlaProValLeu 419  
DB 563 AGTAGAAGCGCTATGAGTCCAGGTAGAGGAGATGTGACTCCAGGGGCCCCAGTACTC 504  
QY 420 ArgValThraIaSerAspArgAspGIySerAsnAlaValIleTySerIleMet 439  
DB 503 CGAGTCAAGCTCTCGATCGAGACAGAGGAGCAATGCCGCGTGCATATGATCATG 444  
QY 440 SerGIyAsnAlaArgGIyGlnPheTyLeuAspAlaGlnThrgIyAlaLeuAspValVal 459  
DB 443 AGTGGCAATGCTCGGGGACAGTTTATCTGAGATCCCACTGAGGCTCTGAGATGTGTG 384  
QY 460 SerProLeuAspTyTrpIleThrgIyGlnTyTrpLeuArgValArgAlaGlnAspGIy 479

```

Db      383 AGCCCTTCTACTATGAGACACACAGAGTACACCTTACGGGGCGACGACAGATGGT 324
Qy      480 GAYAGProProleuSerAanValSerGlyLeuValThrValGlnValLeuAapIleAan 439
Db      323 GGCCTGCCCTCATCTTTATGTCCTGCTTGGTGGACAGACAGGCTCCGAGATATCAAC 264
Qy      500 AspAanAalProIlePheValSerThrProPheGlnAalThrValLeuGlnSerValPro 519
Db      263 GACAAATGCCCATCTTCTGTCAGACACCTTTCCAGGCTACTGTCCTGGAGAGCGTCCC 204
Qy      520 LeuGlyTyrLeuValLeuHsValGlnAalIleAapAalAapAalAapAalAapAalAap 539
Db      203 TTAGGCTACCTGCTTCTCCATGTCACAGGCTATGACCGCTGATCTGCTGCAATGCCCG 144
Qy      540 LeuGlyTyrArgLeuAalGlyValGlyHsAapPheProPheThrIleAanAanGlyThr 559
Db      143 CTGGAAATCCGCTTCTGCTGGAGGAGACATGACTTCCCTTCCATCAACATGAGCA 84
Qy      560 GlyTyrIleSerValAalAalGlnLeuAapAalGlnGlnValAalAapPheTyrSerPheGly 579
Db      83 GGTGGATCTCTGTGGCTGCTGACATGACCGGAGAGAGATTGATTTCTACAGCTTTGG 24
Qy      580 ValGlnAalArgAapPheGly 586
Db      23 GTGAAAGCTCGAGACCATGGC 3

```

```

RESULT 7
BX409741 877 bp mRNA linear EST 13-MAY-2003
LOCUS BX409741 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0D013YE18 5-PRIME, mRNA sequence.
ACCESSION BX409741
VERSION BX409741.1 GI:30652999
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 877)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 64.f For more
information about this cluster, see http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF012ZG07.AF01112.1&cluster=64.f.
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue Genoscope sequence ID: CS0BAF012ZG07.AF01112_1.
Location/Qualifiers

```

```

FEATURES
Source
1..877

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D013YE18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 140 a 315 c 226 g 194 t 2 others
ORIGIN

```

```

Alignment Scores: 3.31e-172 Length: 877
Pred. No.:

```

```

Score: 204.00 Matches: 204
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.98% Indels: 0
DB: 13 Gaps: 0

```

```

US-09-916-849a-3 (1-2923) x BX409741 (1-877)

```

```

Qy      2441 ValIleAalIleLeuHsIlePheLeuTyrLeuGlyThrPheSerThrAalLeuLeuGlu 2460
Db      29 GTGATTTGCCATCTCTGTCGACTTCTGTACTTCTGACCTTTCTTCTGGCTCTGCGAG 88
Qy      2461 AlaLeuHsIleTyrArgAalLeuThrGlnValAalAapValAanThrGlyProMetArg 2480
Db      89 GCTTGACCTGTACCGGGCACTCACTAGAGGCGCGAGTCAACACCGGCCCATGGCC 148
Qy      2481 PheTyrTyrMetLeuGlyTyrGlyValProAalPheIleThrGlyLeuAalValGlyLeu 2500
Db      149 TTCTACTACATGCTGGGCTGGGGCGTGGCTGCTTTCATCAACAGGCTACCGTGGGCTG 208
Qy      2501 AapProGlnGlyTyrGlyAanProAapPheGlyTyrPleuSerIleTyrAapThrLeuIle 2520
Db      209 GACCCGAGGGCTACGGGAACTTCACTTCTGCTGCTGCTCAATATGACAGCTCATC 268
Qy      2521 TTPSerPheAalGlyProValAalPheAalAalSerMetSerValPheLeuTyrIleLeu 2540
Db      269 TCGAGTTTGTGCGCCCGGTGGCTTTGGCCGTCTGATGATGATGTCCTTCTGTACATCTG 328
Qy      2541 AlaAalAalAalSerCysAalAalAalArgGlnGlyPheGlnValGlyGlyProValSer 2560
Db      329 GGGGCCCGGCTCTGCTGGCTGGCCAGCGGAGGCTTGAAGAAAGTCTCTGTCG 388
Qy      2561 GlyLeuGlnProSerPheAalAalLeuLeuLeuSerAalThrTyrPleuAalLeu 2580
Db      389 GGCCTGACAGCTCTTGTGGCGCTCTTCTGCTGAGCGCACCTGCTGCTGGACATG 448
Qy      2581 LeuSerValAanSerAapThrLeuLeuPheHsIleTyrLeuPheAalThrCysAanCysIle 2600
Db      449 CTCTGTGTCACAGAGACACCTCTCTTCCACTCACTTGTGCTACCTGCAATGTCATC 508
Qy      2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGlnValArgValAalLeu 2620
Db      509 CAGGCCCTTATATCTCTCTCTCTATGAGGCTTACAGAGAGTCCGGAAGCATC 568
Qy      2621 LysLeuAalCysSerAryLysProSerProAapProAalLeuThrThrLysSerThrLeu 2640
Db      569 AAGCTTGCTGAGCGCGAGCGGACGCCCTTACCTGTCTGACACCAAGTCCACCTCG 628
Qy      2641 ThrSerSerTyr 2644
Db      629 ACCTGCTCTAC 640

```

```

RESULT 8
BQ677400 906 bp mRNA linear EST 15-JUL-2002
LOCUS BQ677400 AGENCOURT 8209038 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260299
DEFINITION 5', mRNA sequence.
ACCESSION BQ677400
VERSION BQ677400.1 GI:21790079
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (UMNL)

```



Db	183	GCGCTTACCCACTATCATGACAGTGAGGAGAAGAAAGCAAGCAGAGCAAGAGAGAGGCCGCC	242
Oy	2755	PheProgl1yGIuGlInGLYTTTPApSeSerLeuEnGLYProGL1yAlaGluIuaLeuProLeu	2774
Db	243	TYTCCTGGAGAGCAGGAGCTGGAGTAGCTGTGCGGGCCCTGGAGGCAGAGACTGCCCTTG	302
Oy	2775	HISerThrPOLyVAsPGL1yGLYProGL1yProGL1yLysAlaProTrpProGL1yAspPhe	2794
Db	303	CACAGTACTCCCAAGAGATGGGGGCCCAAGGCTGGCAAGCCCCCTGGCCAGAGACTTT	362
Oy	2795	GLYThrThraLYALyGLIsErSerGLYanGL1yAlaProGL1uGLIuaLeuArgLIuan	2814
Db	363	GGGACCACAGAAAAAGAGATAGTGTGCACAAGGGGCCCTCTGAGAGCGGCTCGGGAGAAAT	422
Oy	2815	GLYAspAlaLeu-SerArGGLUGL1SerLeuGL1yProLeuProGL1ySsrSerAlaGlnPr	2834
Db	423	GGAGATGCCCTTNGTCTCCAGAGGGGTCCTTAGGCCCTTCAGGCTCTTCTGCCCAGCC	482
Oy	2834	OHIsLySGl1yILeLeuLYSLySLyCyaleuProThrIIeserGL1yLSerSerLeuLe	2855
Db	483	TCACAAAAGCATCTTAAGAAAGATGTCTGCCCAACATCAGCGAAGAGAGAGAGCTCTCT	542
Oy	2854	UARGLeuProLeuGLInGLInCYethGLYSerSerArgLYserSerAlaSerGLInGLYse	2874
Db	543	GCGGCTCCCCCTGGAGCAATGACAGGGTCYTCCGCGGGGTCTCTCGTAAAGAGGGCAG	602
Oy	2874	rARgLYGLYPro 2878	
Db	603	CCGGGGCGGNCCTG 615	
RESULT 10			
LOCUS	BQ676596	930 bp	mRNA linear EST 15-JUL-2002
DEFINITION	AGENSCOURT 8196790 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6263505		
ACCESSION	BQ676596		
VERSION	BQ676596.1 GI:21789275		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 930)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: rsgr@bbsr@mail.nih.gov Tissue Procurement: DCD/DMP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LRCM2427 row: 1 column: 10 High quality sequence stop: 602. Location/Qualifiers 1..930 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6263505" /tissue_type="melanotic melanoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC_112" /note="Organ: skin; Vector: pOT97; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and		
FEATURES			
source			

Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

BASE COUNT	178 a	353 c	271 g	127 t	1 others					
ORIGIN										
Alignment Scores:										
Pred. No.:	8,27e-158	Length:	930							
Score:	188.00	Matches:	201							
Percent Similarity:	99.50%	Conservative:	0							
Best Local Similarity:	99.50%	Mismatches:	0							
Query Match:	6.43%	Indels:	1							
Db:	13	Gaps:	0							
US-09-916-849A-3 (1-2923) x B0676596 (1-930)										
QY	2675	serlgnproserlyrrilepropheleuarglgnluseralaleuasnproglgln	2694							
Db	3	AGTAGCCCGACGACTACATCCCTCTTCTGAGGAGGATCCGACATCGAACCTCGGCCCA	62							
QY	2695	glYproBroglyleuaglYaspproglyserleuphelleuaglYglnYaspglnghis	2714							
Db	63	GGGCCCCCTGGCCCTGGGGGATCCAGGACGACTGTCTTCTGAAAGGTCAAGACCAAGCAT	122							
QY	2715	AspProAspPhrAspSerAspSerAspLeuSerleuGlunAspAspGlnSerGlySerTyr	2744							
Db	123	GATCTTCAACGCGACTCCGACAGGACCTGTCTCTTAAAGCGACGAGAGGCTCTAT	182							
QY	2735	AlaSerThrHisSerSerAspSerSerGlunGlunGlunGlunGlunGlunGlunAla	2754							
Db	183	GCCCTTACCCACTATACACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	242							
QY	2755	PheProGlyglunGlunGlyTTPAspSerleuGlunGlyProGlyAlaGlunArgleuProleu	2774							
Db	243	TTCCCTGAGAGGACGAGGCTGGGATAGCTGTGAGGCTGAGACAGAGACTGCCCTG	302							
QY	2775	HisSerThrProlyAspAspGlyglYProGlyProGlyYAspAlaProTTPProGlyAspPhe	2794							
Db	303	CACGATCTCCCAAGAT	362							
QY	2795	glYThrThralayrgluserSerGlyYasnGlyAlaProglunGlunArgleuArglunasn	2814							
Db	363	GGGACCAACAGAAAGAGAGTATGTGGCAACGGGGCCCCCTGAGGAGCGGCTCGGAGGAT	422							
QY	2815	GlYAspAlaleuSerArglunGlYSerleuGlYProleuProGlySerSerAlaGlnPr	2834							
Db	423	GGAAATGCCCTGGTCTGAGAGGGGTCTCTTAGGCCCTTCAGAGCTTTTGTGCCAGCC	482							
QY	2834	ohileYglYileuYslyslYsCybleuProThrIleSerGluYsSerSerleu	2854							
Db	483	TTCAAAGAGCATCTTAAAGAAAGATGTCTGCCCAATACAGCAGAGAAAGACACTCTCT	542							
QY	2854	uArgleuProleuGlunGlunCyThrGlySerSerArglySerSerAlaSerGlyYse	2874							
Db	543	GCGGCTCCCCCTGAGGACATGCAAGGGTCTTCCGGGGCTCTCCGCTAGTGAAGGCGAG	602							
QY	2874	Arg 2875								
Db	603	CCGG 606								
RESULT 11										
B0849091 893 bp mRNA linear EST 16-OCT-2002										
LOCUS B0849091										
DEFINITION AGNCOURT 10441717 NIH MGC 109 Homo sapiens cDNA clone										
IMAGE:6597377 5', mRNA sequence.										
ACCESSION B0849091										
VERSION B0849091.1 GI:24034054										
KEYWORDS EST.										
SOURCE Homo sapiens (human)										
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.										
REFERENCE 1 (bases 1 to 893)										
AUTHORS NIH-WGC <a href="http://wgc.nci.nih.gov/">http://wgc.nci.nih.gov/</a> .										

TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: c9abbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLMT at:  
http://image.llnl.gov  
Plate: L10M2828 row: h column: 17  
High quality sequence stop: 556.  
Location/Qualifiers  
1. 893  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAG6:6597377"  
/issue\_type="teratocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH MGC 109"  
/note="Organ: ovary; Vector: pOTB7; Site 1: SCOR1; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into SCOR1/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

BASE COUNT 179 a 268 c 274 g 172 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6,32e-157 Length: 893  
Score: 187.00 Matches: 187  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.40% Indels: 0  
DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BUB49091 (1-893)

QY 1689 leugluPProG1AARgAlaAsnAerG1YAAPTPhIAHIAIG1nleuAla1eugluYala 1708  
DB 2 CTGAGAGCCAGGCGCGCCATATGACGTGACCTGACAGCTGCGACCTGCGAGCC 61

QY 1709 SerG1YGLYProG1YHIAIAIleuSerPheAspTYRGLYGLnAARgAla1eugluY 1728  
DB 62 AGCGGGGGGCGCGCCATCCATTCCTGCTTCGATTATGGGAGAGAGAGAGAGAGG 121

QY 1729 Aa1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLYlePro 1748  
DB 122 AACCTGGGCGCGCGCTGATGCTGACCTGAGCAATAAAGTGGCGGAATACT 181

QY 1749 G1YProAla1eugluYAla1AARg1YPhIAHIAIG1YleuG1nYAla1Ser 1768  
DB 182 GAGCGAGCGCGCGGTGGCGCGCTTCGAGGCGTGTTCAGAGGTGTGGGCTGAGC 241

QY 1769 AAPTPhProG1YGLYAla1AARg1YProSerHIAIG1YGLYleuA1eugluY 1788  
DB 242 GATACCGCGAGAGGGGTTAAAGCTCGATCCAGCCATGGGAGAGAGAGAGAGAG 301

QY 1789 G1nYleuSer1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLY 1808  
DB 302 CAGAGCTGAGCTGCTGACCTTGTGACTCAAAACCGCTGCTGCTAAGCTATTC 361

QY 1809 Ser1AAsp1eugluYleuSer1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLY 1828  
DB 362 AGCAAGAGCTGAG 421

QY 1829 ThrAa1eugluYleuSer1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLY 1848

DB 422 ACTAATGTGTGTGACCTTAACCCGTGTATGACACCAAGCTGTGTGTATCCCGGAGCCAGT 481

QY 1849 AlAProH1eugluYleuSer1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLY 1868  
DB 482 GCGCCCGAGCTATTAAGCTGAG 541

QY 1869 Arg1leuA1eugluYProA1eugluYHIAIAIleuSerPheAsn1eThrValG1YGLY 1875  
DB 542 AGGATTTGACCAAGCTGTGTCC 562

RESULT 12  
AW370778  
LOCUS AW370778 640 bp mRNA linear EST 04-FEB-2000  
DEFINITION QV0-BT0263-101299-072-G04 BT0263 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW370778  
VERSION AW370778.1 GI:6875432  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS HCCP http://www.ludwig.org.br/ORBSTS.  
TITLE The FAPSP/LICR Human Cancer Genome Project  
JOURNAL Unpublished  
COMMENT  
Contact: Stimpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPSP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL:  
http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0-BT0263-  
101299-072-g04&ct=1999-12-10&cl=1  
Seq primer: puc 18 forward  
High quality sequence stop: 640.  
Location/Qualifiers  
1. 640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_id="Br0263"  
/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (0.5. Letters Patent application No. 196  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 122 a 213 c 160 g 145 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7,32e-150 Length: 640  
Score: 179.00 Matches: 207  
Percent Similarity: 99.04% Conservative: 0  
Best Local Similarity: 99.04% Mismatches: 1  
Query Match: 6.12% Indels: 2  
DB: 9 Gaps: 0

US-09-916-849A-3 (1-2923) x AW370778 (1-640)

QY 2247 AlAserVal1leu1eYrAARgThleuAla1Yleu1euproH1AAsnTYRAspProAsp 2266  
DB 15 GCGAGCGATCATCTATCCGACCTGCGCGGCTACTGCTCAATATGAGACCTGAC 74

QY 2267 LysArgSer1eugluYAla1ProLYAARg-Pro1le1AAsnThrProVal1Aser1e 2286  
DB 75 AAGCGAGCTTGAAGTCCCAATG-CGCGATATCAACACCCGCTGGAGCATCAG 133



QY 2286 rValhAspAspGluGluLeuLeuProArgAlaLeuAspLeuProValThrValGlnph 2306  
 Db 134 CGTCCATGATGATGAGGAGCTTCTGCCCCGGGCGCTGCAAAACCGCTGACGTTGACGTT 193  
 QY 2306 eArgLeuLeuGluThrGluGluArgThrIlyProIleCysValPheThrAsnHiser11 2326  
 Db 194 CCGCTCTGCTGAGACAGAGGAGCGGACCAAGCCATCTGTGCTCTGGAACCATTCAMT 253  
 QY 2326 eLeuValSerGlyThrGlyGlyIlyTPSerAlaArgGlyCysGluValValPheArgAsn1 2346  
 Db 254 CCGGTGAGTGGGACAGTGGCTGCTGGCCAGAGGCTGGAAGTCTCTCCGCAATGA 313  
 QY 2346 uSerHisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSe 2366  
 Db 314 GAGCCACGTGAGTGGCCAGTGGCAACCATGACAGAGCTGCTGAGTGTCTGACGACGTTTC 373  
 QY 2366 rArgArgGluAsnGlyGluIlyLeuProLeuIlyThrLeuThrValAlaLeuGlyVal 2386  
 Db 374 TCGCGGAGAAATGGGAGATCCCTGCACTGAAGACATGACATGCTGCTAGGTGT 433  
 QY 2386 lThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeuArgGlyLeuArgSe 2406  
 Db 434 CACTTGGGCTGCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 493  
 QY 2406 rAsnGlnHisGlyIlyLeuArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPh 2426  
 Db 494 CAACCAACACGGGATCGGACGTAACTGACAGCTGCTGGGCTGCTGACGTGCTGCT 553  
 QY 2426 eLeuLeuGlyIlyLeuAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaLeuLeu 2446  
 Db 554 CCTCTGGGATCAACAGGCTGACCTCTTGTGCTGCAAGTCAATGCTGCTGCT 613  
 QY 2446 uHisPheLeuTyrlLeuCysThrPhe 2454  
 Db 614 GCACTTCTGCTGCTGCTGCTGCT 638

RESULT 13  
 BE008566/c 556 bp mRNA linear EST 05-JUN-2000  
 LOCUS BE008566  
 DEFINITION R5-BN0158-190400-031-H09 BN0158 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE008566  
 VERSION BE008566.1 GI:8268799  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 556)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunesei,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J..  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 10737800  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPS/PCR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=1&t=R5-BN0158-190  
 400-031-H09&t3=2000-04-19&t4=1)  
 Seq primer: puc 18 forward

FEATURES  
 source High quality sequence start: 5  
 Location/Qualifiers  
 1..556  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0158"  
 /note="Organ: breast normal; Vector: puc18; Site 1: Smal;  
 Site 2: Sma1; A mini-library was made by cloning products  
 derived from ORSSTs PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 139 a 146 c 196 g 75 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.27e-141 Length: 556  
 Score: 169.00 Matches: 169  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.78% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-916-849A-3 (1-2923) x BE008566 (1-556)  
 QY 2396 PhePheLeuThrLeuLeuArgIlyLeuArgSerAsnGlnHisGlyIlyLeuArgAsnLeu 2415  
 Db 519 TTTCTTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 460  
 QY 2416 ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuGlyIlyLeuAsnGlnAlaAspLeu 2435  
 Db 459 AAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400  
 QY 2436 ProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrlLeuCysThrPheSer 2455  
 Db 399 CTTTTTGGCTGACAGTCAATGCCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 340  
 QY 2456 TrpAlaLeuLeuGlnAlaLeuHisLeuTyrlArgAlaLeuThrGluValArgAspValAen 2475  
 Db 339 TGGGCTCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280  
 QY 2476 ThrGlyProMetArgPheTyrlTyrlMetLeuGlyIlyTrpGlyValProAlaPheIleThrGly 2495  
 Db 279 ACCGGCCCATCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 220  
 QY 2496 LeuAlaValGlyLeuAspProGluGlyIlyTrpGlyAsnProAspPheCysTrpLeuSerIle 2515  
 Db 219 CTAGCGGTGGGCTGCTGACCCCGAGGGCTGACGGAACCTGACCTTCTGCTGCTGCTGCTGCTGCT 160  
 QY 2516 TyrlAspThrLeuIlyTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerVal 2535  
 Db 159 TATGACCGCTATCTGAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100  
 QY 2536 PheLeuTyrlLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnIlyPheGluIly 2555  
 Db 99 TTCTGTATATCTCTGCGCGCGCGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 40  
 QY 2556 TyrlGlyProValSerGlyLeuGlnPro 2564  
 Db 39 AAAGTCTGCTCTGCGGCTGCAAGCC 13

RESULT 14  
 BX110290 523 bp mRNA linear EST 07-FEB-2003  
 LOCUS BX110290  
 DEFINITION BX110290 Soares infant brain INIB Homo sapiens cDNA clone  
 IMAGE998K04177 ; IMAGE:42919, mRNA sequence.  
 ACCESSION BX110290  
 VERSION BX110290.1 GI:27836294  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 523)  
 Ebert, L., Heil, O., Hennig, S., Neubert, F., Partsch, E., Peters, M.,  
 Radetof, U., Schneider, D. and Korn, B.  
 TITLE Human Unigeneset - RZPD3  
 JOURNAL Unpublished  
 COMMENT Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
 RZPD: IMAGP998K04177.  
 RZPDLIB: I.M.A.G.E. CDNA Clone Collection;  
 Human Unigeneset - RZPD3 (RZPDLIB No. 972)  
 http://www.rzpd.de/Cloncards/csl-  
 bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de

FEATURES  
 SOURCE This clone is available royalty-free from RZPD;  
 contact RZPD (clone@rzpd.de) for further information. Seq primer:  
 M13u, Primer sequence: GGTGTAAAGCAGCGCCAGT.  
 Location/Qualifiers  
 1..523

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGP998K04177 ; IMAGE:42919"  
 /sex="female"  
 /dev\_stage="73 days post natal"  
 /lab\_host="PH10B (ampicillin resistant)"  
 /clone\_lib="Soares infant brain INTB"  
 /note="Organ: whole brain; Vector: latmid BA; Site 1: Not  
 1; Site 2: Hind III; 1st strand cDNA was primed with a Not  
 1 - oligo(dT) primer [5',  
 AACTGAGAAATTCGGCGCGGAGGATTTTCTTTTCTTTT 3'];  
 double-stranded cDNA was ligated to Hind III adaptors  
 (pharmacia), digested with Not I and directionally cloned  
 into the Not I and Hind III sites of the latmid BA vector.  
 Library went through one round of normalisation. Library  
 constructed by Bento Soares and M.Factina Bonaldo."  
 BASE COUNT 109 a 183 c 133 g 97 t 1 others

Alignment Scores:  
 Pred. No.: 2,95e-138 Length: 523  
 Score: 166.00 Matches: 166  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.68% Indels: 0  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BK110290 (1-523)

QY 2585 SerAapThrLeuPheHisTyrLeuPheAlaThrCysAsnCysAlleglnGlyProPhe 2604  
 DB 26 AAGGACACACCTCTCTTCACTACCTCTTGTCTACTGCAATTCGATCCAGGCGCCCTTC 85  
 QY 2605 IlePheLeuSerTyrValValLeuSerIyrgIValArgIValAlaLeuLysLeuAlaCys 2624  
 DB 86 ATCTTCCTCTCTTAATGTGTCTAGACAGAGGTCCGAAAGCACTCAAGCTTCGCTGC 145  
 QY 2625 SerATgIyPProSerProAlaLeuThrThrIySerThrLeuThrSerSerTyr 2644  
 DB 146 ACCCGAAGCCAGCCCTGCTGTGACCAACCAATCACTGACCTGCTGCTTAC 205  
 QY 2645 AAnCysProSerProTyrAlaApGlyArgLeuTyrGlnProTyrGlyAspSerAlaGly 2664  
 DB 206 AACTGCCCAAGCCCTTACCAAGATGGCGGCTGTACCAAGCCCTTACCAAGATCTCGCGCGC 265

QY 2665 SerLeuHisSerThrSerArgSerGlyIySerGlnProSerTyrIlePheLeuLeu 2684  
 DB 266 TCTTCGACAGCAGCAGCTGCTCGGCGCAAGTCAAGCCAGTACATCCCTTCTTGCCTG 325  
 QY 2685 ArgGlnIySerAlaLeuAsnProGlyGlnGlyProPheGlyLeuGlyAspProGlySer 2704  
 DB 326 AAGGAGAGTCCGACCTGACCTGAGGAGGCGCCCTGCGCTGGGGATCCAGCGAGC 385  
 QY 2705 LeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeu 2724  
 DB 386 CTGTTCTCGAAGAGTCAGACAGCAGCATGATCTTGAACAGGACTCCGACGATCACTCG 445  
 QY 2725 SerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlu 2744  
 DB 446 TCTTGAAGAGCAGCAGCAGATGCTCTTATGCTCTTCACTCACTCACTCACTCACTGAGG 505  
 QY 2745 GlnGlnIyGlnGlnGlnGlu 2750  
 DB 506 GAAAGAGAGAGAGAGAA 523

RESULT 15  
 BF951925/c 549 bp mRNA linear EST 22-JAN-2001  
 LOCUS QY1-NN0228-091100-440-f05 NN0228 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BF951925  
 ACCESSION BF951925.1 GI:12369200  
 VERSION BF951925.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 549)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsumura, A., Bata, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.V. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 JOURNAL 20202663  
 MEDLINE 10737800  
 PUBMED Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=QY1-QY1228-QY1-NN0228-  
 091100-440-f05&f3=2000-11-09&f4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 4  
 High quality sequence stop: 549.  
 Location/Qualifiers  
 1..549

# FEATURES

## SOURCE

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="NN0228"  
 /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI,  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORSTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 113 a 173 c 158 g 105 t

## ORIGIN

## Alignment Scores:

Pred. No.:	2,44e-137	Length:	549
Score:	165.00	Matches:	165
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.64%	Indels:	0
DB:	10	Gaps:	0

US-09-916-849a-3 (1-2923) x BP951925 (1-549)

```

QY      1707  GYAlaSerGlyProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAla 1726
           |||
DB      547  GAGGCCAGGGGGGGGGCTGGCCATGCCATTCTCTCGATTATGGGACGAGAGACA 488
           |||
QY      1727  GUGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyGly 1746
           |||
DB      487  GAGGGCAACCTGGGGCCCCCGGCTGCATGCTCTGCACCTGCAGCAATAAACAGTGGGGCGGA 428
           |||
QY      1747  IleProGlyProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArg 1766
           |||
DB      427  ATACCTGGGCGACCGCGGGGTGGCCGCGCTTCGGGGCTGTTGCAAGGTGTGCGG 368
           |||
QY      1767  ValSerAspThrProGlnGlyValAsnSerLeuAspProSerHisGlyGlnSerIleAsn 1786
           |||
DB      367  GTGAGCGATACCGCGAGGGGGTTAACAGCTCGATCCAGCCATGGGGGAGAGCATCAAC 308
           |||
QY      1787  ValGlnGlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSer 1806
           |||
DB      307  GTGGAGCAAGGCTGTAGCTGCTGACCCCTGTGACTCAAAACCGTGTCTCTGCTTAACAGC 248
           |||
QY      1807  TyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAsp 1826
           |||
DB      247  TATTGACGACACACACTGGGACACTATTCTTGCACTGTGATCAAGTTACTATGGTGAC 188
           |||
QY      1827  AsnCysThrAsnValCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLys 1846
           |||
DB      187  AACTGTACTATATGTGTGACCTGAACCCGTGTAGAGACCAAGCTGTGTGTACCCGCAAG 128
           |||
QY      1847  ProSerAlaProHisGlyTyrThrCysGlnCysProProAsnTyrLeuGlnProTyrCys 1866
           |||
DB      127  CCAAGTGCCCCCATGGCTATACCTGGAGTGTCCCCCAATTACCTGGGGCCATACTGT 68
           |||
QY      1867  GluThrArgIleAsp 1871
           |||
DB      67  GAGACCGAGATTGAC 53
           |||

```

Search completed: February 14, 2004, 21:11:24  
 Job time : 11573 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:57:15 ; Search time 66 Seconds

(Without alignments)  
9273.097 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 15545  
Sequence: 1 MRSPTGVLPPTPPPLLL.....AGTVDESSGSEFLFNFLLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15545	100.0	2923	9 US-09-788-711A-4	Sequence 4, Appli
2	15545	100.0	2923	11 US-09-916-849A-3	Sequence 3, Appli
3	15545	100.0	2923	12 US-10-174-677-29	Sequence 29, Appli
4	15545	100.0	2923	12 US-10-120-801-53	Sequence 53, Appli
5	15545	100.0	2923	12 US-10-292-798-932	Sequence 92, Appli
6	15545	100.0	2923	15 US-10-225-567A-524	Sequence 524, App
7	15518.5	99.8	2956	9 US-09-788-711A-2	Sequence 2, Appli
8	8974.5	57.7	3034	9 US-09-737-149-25	Sequence 25, Appli
9	8974.5	57.7	3034	9 US-09-737-149-30	Sequence 30, Appli
10	8974.5	57.7	3034	12 US-10-131-409-70	Sequence 70, Appli
11	8974.5	57.7	3034	12 US-10-120-801-52	Sequence 52, Appli
12	8974.5	57.7	3034	12 US-10-150-811-70	Sequence 70, Appli
13	8754	56.3	3014	9 US-09-737-149-2	Sequence 2, Appli
14	8754	56.3	3014	12 US-10-241-230-107	Sequence 107, Appli
15	8754	56.3	3014	12 US-10-174-677-77	Sequence 77, Appli

16	8754	56.3	3014	12 US-10-295-027-750	Sequence 750, App
17	8754	56.3	3014	12 US-10-240-145-63	Sequence 43, Appli
18	8754	56.3	3014	15 US-10-225-567A-444	Sequence 444, App
19	8709	56.0	3028	12 US-10-131-409-14	Sequence 14, Appli
20	8709	56.0	3028	12 US-10-131-409-69	Sequence 69, Appli
21	8709	56.0	3028	12 US-10-150-811-14	Sequence 14, Appli
22	8709	56.0	3028	12 US-10-150-811-69	Sequence 69, Appli
23	7833.5	50.4	3312	15 US-10-225-567A-656	Sequence 656, App
24	7780.5	50.1	3313	9 US-09-737-149-29	Sequence 29, Appli
25	7763	49.9	3298	15 US-10-149-819-21	Sequence 21, Appli
26	4951	31.8	1713	9 US-09-737-149-27	Sequence 27, Appli
27	4817.5	31.0	3579	12 US-10-174-677-5	Sequence 5, Appli
28	4817.5	31.0	3606	12 US-10-120-801-54	Sequence 54, Appli
29	3641	23.4	717	10 US-09-925-300-1299	Sequence 1299, Ap
30	2958	19.0	568	9 US-09-843-856-2	Sequence 2, Appli
31	2937	18.9	568	12 US-10-264-237-2041	Sequence 2041, Ap
32	2900.5	18.7	565	12 US-10-176-847-100	Sequence 100, App
33	2491	16.0	2610	12 US-10-369-493-6553	Sequence 6553, Ap
34	2491	16.0	2610	12 US-10-369-493-6554	Sequence 6554, Ap
35	2448.5	15.8	646	12 US-10-017-161-1096	Sequence 1096, Ap
36	1712.5	11.0	4961	12 US-10-114-153-64	Sequence 64, Appli
37	1534.5	9.9	5147	12 US-10-174-677-4	Sequence 4, Appli
38	1516.5	9.8	4590	12 US-09-970-944-21	Sequence 21, Appli
39	1516.5	9.8	4590	12 US-10-295-027-1323	Sequence 1323, Ap
40	1516.5	9.8	4590	15 US-10-160-758-13	Sequence 13, Appli
41	1516.5	9.8	4590	15 US-10-160-758-14	Sequence 14, Appli
42	1516.5	9.8	4590	15 US-10-060-036-157	Sequence 157, App
43	1460	9.4	458	12 US-10-029-386-33121	Sequence 33121, A
44	1419.5	9.1	3097	12 US-10-174-677-6	Sequence 6, Appli
45	1402	9.0	4544	12 US-10-085-198-22	Sequence 22, Appli

## ALIGNMENTS

RESULT 1  
US-09-788-711A-4  
; Sequence 4, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamara Testa  
; TITLE OR INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788, 711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-4

Query Match 100.0%; Score 15545; DB 9; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRSPTGVLPPTPPPLLLLLPPPLADGVPCRSLSRRGSSGACAPMGWLCPS	60
DB	1	MRSPTGVLPPTPPPLLLLLPPPLADGVPCRSLSRRGSSGACAPMGWLCPS	60
QY	61	SASINMTYSCRDRAGTETLNGHLPVHDGLRVCPSEBAHPLPAPBGCWSCRLLGIG	120
DB	61	SASINMTYSCRDRAGTETLNGHLPVHDGLRVCPSEBAHPLPAPBGCWSCRLLGIG	120
QY	121	GHSPOGKTLTPEHPCIKAPRLRCQSCIKLAQAPGLRAGERSPEESIGRRKRVNTAPQ	180
DB	121	GHSPOGKTLTPEHPCIKAPRLRCQSCIKLAQAPGLRAGERSPEESIGRRKRVNTAPQ	180
QY	181	FOPSYQNTVPENDPATTPVASLPAIDPDGEAGRLTYTMDALPDSNSNOFSIDPVTGA	240
DB	181	FOPSYQNTVPENDPATTPVASLPAIDPDGEAGRLTYTMDALPDSNSNOFSIDPVTGA	240

Db 181 POPSYOATVBNQAPGTPVASLRAIDPDGEGARLEYTMALFDSRSNOQFSLDEVTGA 240  
QY 241 VTTAAEIDRETKSTHFRVTAODHGMRRSALATLTITVTDNDHDVPFEGQCKESLAE 300  
Db 241 VTTAAEIDRETKSTHFRVTAODHGMRRSALATLTITVTDNDHDVVEEQCKESLAE 300  
QY 301 NLEVGVEVLTVRATDGDAPPNANILVRLLEGSGSGSPSEVEIDPRSGVIRTRGPVDBREV 360  
Db 301 NLEVGVEVLTVRATDGDAPPNANILVRLLEGSGSGSPSEVEIDPRSGVIRTRGPVDBREV 360  
QY 361 ESQVLTVEASDQGRDPRSTTAAPLSTVEDDNDNAQPSBKRYVYQVEDVTGAPVLR 420  
Db 361 ESQVLTVEASDQGRDPRSTTAAPLSTVEDDNDNAQPSBKRYVYQVEDVTGAPVLR 420  
QY 421 VTSASDDBKGSNAVHYSIMSGNARGOFLDAOTGALDVSPPLDYETTKETTLKVRADGG 480  
Db 421 VTSASDDBKGSNAVHYSIMSGNARGOFLDAOTGALDVSPPLDYETTKETTLKVRADGG 480  
QY 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLSBVPJGYLVLAHQAIDADAGNARL 540  
Db 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLSBVPJGYLVLAHQAIDADAGNARL 540  
QY 541 EYRLAGVGHDPPTTINNGTGMISVAALDREVDFTSFGYBARDHGTPTALTAASVTV 600  
Db 541 EYRLAGVGHDPPTTINNGTGMISVAALDREVDFTSFGYBARDHGTPTALTAASVTV 600  
QY 601 LDVNDNNPFTQEPETVRLMEDAVGTSVTVSAVDHDSVITVQTSQNTNRPSITS 660  
Db 601 LDVNDNNPFTQEPETVRLMEDAVGTSVTVSAVDHDSVITVQTSQNTNRPSITS 660  
QY 661 OSQGGVLSIALPLDYKLERQYVLAVTASDGTRODTAQIVNVTDANTHRPVPQSSHYTV 720  
Db 661 OSQGGVLSIALPLDYKLERQYVLAVTASDGTRODTAQIVNVTDANTHRPVPQSSHYTV 720  
QY 721 VNEDRPAQTIVYLSATDEPTGENARITYEMEDSIPOFRIDAPTGAVTQAEIDYDQVS 780  
Db 721 VNEDRPAQTIVYLSATDEPTGENARITYEMEDSIPOFRIDAPTGAVTQAEIDYDQVS 780  
QY 781 YTLAITARNDNGIPQKSDPTTYLELIVNDVNDNAPOPLADSQSGSYEDVPPFTSVLQISAT 840  
Db 781 YTLAITARNDNGIPQKSDPTTYLELIVNDVNDNAPOPLADSQSGSYEDVPPFTSVLQISAT 840  
QY 841 DROSGLANGRYFTTQGGDDGDGDFIVASTSGIYRTLRRLRENVAYQVLRAYVNDKMP 900  
Db 841 DROSGLANGRYFTTQGGDDGDGDFIVASTSGIYRTLRRLRENVAYQVLRAYVNDKMP 900  
QY 901 ARTPEMETVTVLVNDNDPVPFBOEDPVPFVENSPIGLAVARVATADPDEGTNAQIMYOI 960  
Db 901 ARTPEMETVTVLVNDNDPVPFBOEDPVPFVENSPIGLAVARVATADPDEGTNAQIMYOI 960  
QY 961 VEGNIPPEVQOLDIFSGELTALVDLDYEDREPEYLVITQATSAPLVSRATVHRLDRNDP 1020  
Db 961 VEGNIPPEVQOLDIFSGELTALVDLDYEDREPEYLVITQATSAPLVSRATVHRLDRNDP 1020  
QY 1021 PVYGNPEIILNNYVYTNRSSPFGGATGRVPAHDPDISLITYSPERGENEISVTLNASTG 1080  
Db 1021 PVYGNPEIILNNYVYTNRSSPFGGATGRVPAHDPDISLITYSPERGENEISVTLNASTG 1080  
QY 1081 ELKLSRALDNNRPLEAIMSVLVSQGVSTAOACALRTIITDEMLTSTILRLIEDNSPER 1140  
Db 1081 ELKLSRALDNNRPLEAIMSVLVSQGVSTAOACALRTIITDEMLTSTILRLIEDNSPER 1140  
QY 1141 FLSPBLGLFTIOAAVATLATPPDHVVFNVRDTPDAPGHILNVSLSVGOPPGGGBPFL 1200  
Db 1141 FLSPBLGLFTIOAAVATLATPPDHVVFNVRDTPDAPGHILNVSLSVGOPPGGGBPFL 1200  
QY 1201 PSEBLOERLYNBSLTLAISAGRVLPBDNCLAREPENTMRCVSVRPSSAPFLASS 1260  
Db 1201 PSEBLOERLYNBSLTLAISAGRVLPBDNCLAREPENTMRCVSVRPSSAPFLASS 1260  
QY 1261 VLFRRPIHPVGGALRCRCPGFTGDYCFTEVDLCYSRPGPHGRCSRGGYTCLCRDGYTG 1320  
Db 1261 VLFRRPIHPVGGALRCRCPGFTGDYCFTEVDLCYSRPGPHGRCSRGGYTCLCRDGYTG 1320

QY 1321 BECEVSABSGRCTPGVCKNGGCVNLVYGFKCDPCSGDFEKPVCQVTRSPFASFTTF 1380  
Db 1321 BECEVSABSGRCTPGVCKNGGCVNLVYGFKCDPCSGDFEKPVCQVTRSPFASFTTF 1380  
QY 1381 RGLRORFHTTLSPATYERDGLLYNGRFNEKHDVLALEVIQBOVOLTFSAGESTTVS 1440  
Db 1381 RGLRORFHTTLSPATYERDGLLYNGRFNEKHDVLALEVIQBOVOLTFSAGESTTVS 1440  
QY 1441 PVPVGSVDGQWHTVQLYKYNPVLGQGTLPQGBSEQKAVAVTVNDGCDGVLRFRSVTG 1500  
Db 1441 PVPVGSVDGQWHTVQLYKYNPVLGQGTLPQGBSEQKAVAVTVNDGCDGVLRFRSVTG 1500  
QY 1501 NTSCAAQGTQGGSKSLDTGLPBLIGVBDLPESPEVPMRQPFVGCGRNLQVDSRHIDMAD 1560  
Db 1501 NTSCAAQGTQGGSKSLDTGLPBLIGVBDLPESPEVPMRQPFVGCGRNLQVDSRHIDMAD 1560  
QY 1561 FIANNGTVPQCPAKNOVDSNTCHNGGTCVNQMDAPSCBCEPLGFGKSCAQOMANPOHFL 1620  
Db 1561 FIANNGTVPQCPAKNOVDSNTCHNGGTCVNQMDAPSCBCEPLGFGKSCAQOMANPOHFL 1620  
QY 1621 GSSLVAMHGLSLPISQPMYLSMFRTRQADGVLAQITRGSTITLQAREGHVMTSVBGT 1680  
Db 1621 GSSLVAMHGLSLPISQPMYLSMFRTRQADGVLAQITRGSTITLQAREGHVMTSVBGT 1680  
QY 1681 GLQASSLRLEPRANDGDMHHAQLALGASGGGHAHLSFDYQOQBAEGNLGRLHGLHS 1740  
Db 1681 GLQASSLRLEPRANDGDMHHAQLALGASGGGHAHLSFDYQOQBAEGNLGRLHGLHS 1740  
QY 1741 NITVGGIPGPAGVARGFRGCIQGVRSVDPBGVNSLDBSHGESINVEQCSLPDPDCSN 1800  
Db 1741 NITVGGIPGPAGVARGFRGCIQGVRSVDPBGVNSLDBSHGESINVEQCSLPDPDCSN 1800  
QY 1801 PCPANSYCSNMDYSQSCDPEYGDNCNVCNVDLPCEHOSQCTKPSAPPHGYTCECPN 1860  
Db 1801 PCPANSYCSNMDYSQSCDPEYGDNCNVCNVDLPCEHOSQCTKPSAPPHGYTCECPN 1860  
QY 1861 YLGPYCESTRIDPCRGMMWGHPTGPCNCDVSKGFPDPCNKTSGECHKENHYRPPGSP 1920  
Db 1861 YLGPYCESTRIDPCRGMMWGHPTGPCNCDVSKGFPDPCNKTSGECHKENHYRPPGSP 1920  
QY 1921 CILCOCYPTGSLSRVCDPDEGQCKPGVITGQPCORCONPFAVYTNNGEYVYDSCPRAI 1980  
Db 1921 CILCOCYPTGSLSRVCDPDEGQCKPGVITGQPCORCONPFAVYTNNGEYVYDSCPRAI 1980  
QY 1981 EAGIMWPRTRFGLPAAPCPKSGFTAVARHCDHRGMLPPLNFCTSLTFSEILKGAERL 2040  
Db 1981 EAGIMWPRTRFGLPAAPCPKSGFTAVARHCDHRGMLPPLNFCTSLTFSEILKGAERL 2040  
QY 2041 QNRESGLDSGRSQOLALLRNATQHTAGYFSGDVAVAYQALATRLLAHESDORFGLSATQ 2100  
Db 2041 QNRESGLDSGRSQOLALLRNATQHTAGYFSGDVAVAYQALATRLLAHESDORFGLSATQ 2100  
QY 2101 DVHFTENILRVGSALLDPANKRMHMLIOQTBGCTAMLOHYEYAVASALANNRHMYLSPF 2160  
Db 2101 DVHFTENILRVGSALLDPANKRMHMLIOQTBGCTAMLOHYEYAVASALANNRHMYLSPF 2160  
QY 2161 TIVTPNIVISVVRLLKGNFAGAKLPRYEALRGBOPPDLTETVILPESVRETPPVVRPAG 2220  
Db 2161 TIVTPNIVISVVRLLKGNFAGAKLPRYEALRGBOPPDLTETVILPESVRETPPVVRPAG 2220  
QY 2221 PGBAOPEERLARORRHPELSGGBAVASYITRTLAGLLPHNYDDPKSLARPAPRIINT 2280  
Db 2221 PGBAOPEERLARORRHPELSGGBAVASYITRTLAGLLPHNYDDPKSLARPAPRIINT 2280  
QY 2281 PIVYSISVHDEELPRALDKPVTVOFRLLETEBRTKPCIVFNHNSILVSGGMSARGCE 2340  
Db 2281 PIVYSISVHDEELPRALDKPVTVOFRLLETEBRTKPCIVFNHNSILVSGGMSARGCE 2340  
QY 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKLTLYVALGVTLLAALLTFEFLTL 2400  
Db 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKLTLYVALGVTLLAALLTFEFLTL 2400

QY 2401 LRIASNGHTRNLTAALGLAQLVFLGINDADIPACTVYAIILHFLYCTESWALLE 2460  
 DB 2401 LRIASNGHTRNLTAALGLAQLVFLGINDADIPACTVYAIILHFLYCTESWALLE 2460  
 QY 2461 ALHLRYALTEVDVNTGPMRFYMLGMGVPAPETITGAUGLDEGYNDFCMLSYDITLI 2520  
 DB 2461 ALHLRYALTEVDVNTGPMRFYMLGMGVPAPETITGAUGLDEGYNDFCMLSYDITLI 2520  
 QY 2521 WSPAGPVAFAVMSVFLYTLAARASCAAROGFEKKGPVSGLOPFAVILLISATWLLAL 2580  
 DB 2521 WSPAGPVAFAVMSVFLYTLAARASCAAROGFEKKGPVSGLOPFAVILLISATWLLAL 2580  
 QY 2581 LSVNSDTLFFHFLFKTNCICIQPFIFLSVYVLSKVRKALKLACRKSPPDALTTKSTL 2640  
 DB 2581 LSVNSDTLFFHFLFKTNCICIQPFIFLSVYVLSKVRKALKLACRKSPPDALTTKSTL 2640  
 QY 2641 TSSVNCSPYADGRLYKOPYGSAGSLHSTRSGKSQPSYIIFLREESALNPGQPPGLG 2700  
 DB 2641 TSSVNCSPYADGRLYKOPYGSAGSLHSTRSGKSQPSYIIFLREESALNPGQPPGLG 2700  
 QY 2701 DPGSLFLEGQDQHDPTDSDSLSLDDQSGSVASTHSDSEEBEEREAAPPGQ 2760  
 DB 2701 DPGSLFLEGQDQHDPTDSDSLSLDDQSGSVASTHSDSEEBEEREAAPPGQ 2760  
 QY 2761 WDSLIGPBAERLPLHSTROGPGPKAPMPDPCGTAKESGNAPEERLRENDALSR 2820  
 DB 2761 WDSLIGPBAERLPLHSTROGPGPKAPMPDPCGTAKESGNAPEERLRENDALSR 2820  
 QY 2821 EGSILGFLGSSAOPKIKKCLPTISEKSLPLPEOCTGSSRGSASBGRSGPP 2880  
 DB 2821 EGSILGFLGSSAOPKIKKCLPTISEKSLPLPEOCTGSSRGSASBGRSGPP 2880  
 QY 2881 RPPRQSLQEBQNGVMPIAMSIKAGTVDEDSGSEFLFNFLLH 2923  
 DB 2881 RPPRQSLQEBQNGVMPIAMSIKAGTVDEDSGSEFLFNFLLH 2923

RESULT 2  
 US-09-916-849a-3  
 ; Sequence 3, Application US/09916849A  
 ; Publication No. US20030086934A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bostein, et al.  
 ; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents  
 ; TITLE OF INVENTION: Uses thereof  
 ; FILE REFERENCE: 2002850-0024  
 ; CURRENT APPLICATION NUMBER: US/09/916, 849A  
 ; CURRENT FILING DATE: 2001-07-26  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 3  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Cadherin EGF  
 ; OTHER INFORMATION: LAG Seven Pass G-Type Receptor 2  
 US-09-916-849a-3

Query Match 100.0%; Score 15545; DB 11; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRSPTATGVPPLPPPLLLLLLLPPPLGQVPCRSLSRGRSSGACAPMGWLCPS 60  
 DB 1 MRSPTATGVPPLPPPLLLLLLLPPPLGQVPCRSLSRGRSSGACAPMGWLCPS 60  
 QY 61 SASNMLTYSRCRDGTELTGHLVPHNGLRWYWCSESAHPLPPAPGCTWSCLLATIG 120  
 DB 61 SASNMLTYSRCRDGTELTGHLVPHNGLRWYWCSESAHPLPPAPGCTWSCLLATIG 120  
 QY 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRKRKNVNTAPQ 180  
 DB 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRKRKNVNTAPQ 180

DB 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRKRKNVNTAPQ 180  
 QY 181 FQPSYQATVPENOPAGTPVVASLRAIDPDEGEAGRLLEYMDALDPSRNSQFSLDPYGA 240  
 DB 181 FQPSYQATVPENOPAGTPVVASLRAIDPDEGEAGRLLEYMDALDPSRNSQFSLDPYGA 240  
 QY 241 VTTAEELDRETKSTHVRVTAQDHGMRRSALATLTTLVTDNDHDPVFEQOEKESLRE 300  
 DB 241 VTTAEELDRETKSTHVRVTAQDHGMRRSALATLTTLVTDNDHDPVFEQOEKESLRE 300  
 QY 301 NLEFYGVLVTRADGAPAPNANTLYRLGSSGSSPEVEPIDRSGVITRRGVDRBEV 360  
 DB 301 NLEFYGVLVTRADGAPAPNANTLYRLGSSGSSPEVEPIDRSGVITRRGVDRBEV 360  
 QY 361 ESYOLTVEASDOGRDPPRSTTAAVFLSVEDDNDNAQFSEKRYVGVREDVTGAPYLR 420  
 DB 361 ESYOLTVEASDOGRDPPRSTTAAVFLSVEDDNDNAQFSEKRYVGVREDVTGAPYLR 420  
 QY 421 VTTASDRKGSNAVHVSIMGNNARQFYLDAGTALDVSPLDYETTKETTLRVRAODG 480  
 DB 421 VTTASDRKGSNAVHVSIMGNNARQFYLDAGTALDVSPLDYETTKETTLRVRAODG 480  
 QY 481 RPLSNVSGLVTVQVLDINDNAPFVSTPFOATVLESVPGLVLAHQALDADGDNARL 540  
 DB 481 RPLSNVSGLVTVQVLDINDNAPFVSTPFOATVLESVPGLVLAHQALDADGDNARL 540  
 QY 541 EYRLAGVGHDPFPTINNGTGMISVAELDREEDVYFSGYEARDHGTALTAASVSVTV 600  
 DB 541 EYRLAGVGHDPFPTINNGTGMISVAELDREEDVYFSGYEARDHGTALTAASVSVTV 600  
 QY 601 LDVNDNPPFTQPEYTYRLMEDAAVGTSVTVSAVDDAHSVITYQITSGNTRNRSITS 660  
 DB 601 LDVNDNPPFTQPEYTYRLMEDAAVGTSVTVSAVDDAHSVITYQITSGNTRNRSITS 660  
 QY 661 OSGGGLVSLALPLDYKLEROYLAVTASDGTRODTAIVNVNDANHRPVEQSSHVTN 720  
 DB 661 OSGGGLVSLALPLDYKLEROYLAVTASDGTRODTAIVNVNDANHRPVEQSSHVTN 720  
 QY 721 VNEBRPAGTTVVLISATDEDTGENARITVEMEDSIPOFRIDADGAVTTOAELEDEDOVS 780  
 DB 721 VNEBRPAGTTVVLISATDEDTGENARITVEMEDSIPOFRIDADGAVTTOAELEDEDOVS 780  
 QY 781 YTLAITARDNGIQKSTTYTLELIVNDVNDNAQPLFDSIQSGSYEEDVPPTSTVLQISAT 840  
 DB 781 YTLAITARDNGIQKSTTYTLELIVNDVNDNAQPLFDSIQSGSYEEDVPPTSTVLQISAT 840  
 QY 841 DRDSGLNGRVFTYFQGGDDGDPFVVESTGIVKTLRLRENVAAQVLAAYADKMP 900  
 DB 841 DRDSGLNGRVFTYFQGGDDGDPFVVESTGIVKTLRLRENVAAQVLAAYADKMP 900  
 QY 901 ARTPEVTVTVLVNDNPPVFEQDEPDVFEENSPIGLAAVARVATDPDEGTNAQIMYQI 960  
 DB 901 ARTPEVTVTVLVNDNPPVFEQDEPDVFEENSPIGLAAVARVATDPDEGTNAQIMYQI 960  
 QY 961 VEGNIPPEVFDLTFSGELTALVDLDYEDREBYLVIQATSAPLVSRAVTVHRLDRNDP 1020  
 DB 961 VEGNIPPEVFDLTFSGELTALVDLDYEDREBYLVIQATSAPLVSRAVTVHRLDRNDP 1020  
 QY 1021 PVYGNPEILFNNVYVTRNSSFPGCAIGRVAPHDPDISLTVSEPERNEISIVLYANSTG 1080  
 DB 1021 PVYGNPEILFNNVYVTRNSSFPGCAIGRVAPHDPDISLTVSEPERNEISIVLYANSTG 1080  
 QY 1081 ELKLSRALDNNRPLEAIMSVLVSDGVSVTAQCALRYTITDEMLTHTSITLRLEDSMPER 1140  
 DB 1081 ELKLSRALDNNRPLEAIMSVLVSDGVSVTAQCALRYTITDEMLTHTSITLRLEDSMPER 1140  
 QY 1141 FLSPFLGLFTQAAVATATPPDHVYVNVNQRDTPADPGHILANTSLVSGQPPGCGGPPFL 1200  
 DB 1141 FLSPFLGLFTQAAVATATPPDHVYVNVNQRDTPADPGHILANTSLVSGQPPGCGGPPFL 1200  
 QY 1201 PSEDLQERLYNRSILTAISAQVLPDPDNI CLREPENYRCVSVLPFSSAPFIASSS 1260  
 DB 1201 PSEDLQERLYNRSILTAISAQVLPDPDNI CLREPENYRCVSVLPFSSAPFIASSS 1260



QY 1261 VLFRRPHVQGLRCRCPGFTGDCYCEYVDLCYSRCPGPHGRCSRREGGVTCICRDGYG 1320  
 DB 1261 VLFRRPHVQGLRCRCPGFTGDCYCEYVDLCYSRCPGPHGRCSRREGGVTCICRDGYG 1320  
 QY 1321 EHEVBARSGRCCTPGVCKNGCTVNLVGGFKDCDPSGDEKPYCOVTRSPRASHPTIP 1380  
 DB 1321 EHEVBARSGRCCTPGVCKNGCTVNLVGGFKDCDPSGDEKPYCOVTRSPRASHPTIP 1380  
 QY 1381 RGLRSHFHTLALSPATKRDGLLLNNGRPNKHDPALEVOEVOULTFSAESTTYS 1440  
 DB 1381 RGLRSHFHTLALSPATKRDGLLLNNGRPNKHDPALEVOEVOULTFSAESTTYS 1440  
 QY 1441 PFWGVGSDGOMHTVQLKTYNKEPLGQGTLPQSPSEQKVAVVVVDGCDTGVALRFGSVLG 1500  
 DB 1441 PFWGVGSDGOMHTVQLKTYNKEPLGQGTLPQSPSEQKVAVVVVDGCDTGVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQSGSKSLDTLGPILLGVPPLPESFPVRMRQFVGCNRNLOVDSRHIDMAD 1560  
 DB 1501 NYSCAAQGTQSGSKSLDTLGPILLGVPPLPESFPVRMRQFVGCNRNLOVDSRHIDMAD 1560  
 QY 1561 FIANNQTPGCPAKKAVCDNTCHNGGTCVNDADAFSCBCLGFGGKSCAOEAMNDPHE 1620  
 DB 1561 FIANNQTPGCPAKKAVCDNTCHNGGTCVNDADAFSCBCLGFGGKSCAOEAMNDPHE 1620  
 QY 1621 GSSLVAMHGLSLPISQPWYLSLMFRTROADGVLLQAITRGRSTITLQLEHGHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISQPWYLSLMFRTROADGVLLQAITRGRSTITLQLEHGHVMSVEGT 1680  
 QY 1681 GLQASSLRLEPGANGDHHAQALGASGPGHALLSPYGOORABGNLGPRLHGLHLS 1740  
 DB 1681 GLQASSLRLEPGANGDHHAQALGASGPGHALLSPYGOORABGNLGPRLHGLHLS 1740  
 QY 1741 NITVGGIPGAGVARGFRCGLQGVNVDTPREGVNSLDPSHGSIINVOGCSLPDCCSN 1800  
 DB 1741 NITVGGIPGAGVARGFRCGLQGVNVDTPREGVNSLDPSHGSIINVOGCSLPDCCSN 1800  
 QY 1801 PCPANSYCSNDMSYSCSDPGYGGDNCTNVCDLNPCEHQSVCTRKPSAHGYTCCEPPN 1860  
 DB 1801 PCPANSYCSNDMSYSCSDPGYGGDNCTNVCDLNPCEHQSVCTRKPSAHGYTCCEPPN 1860  
 QY 1861 YLGPYCESTRIDOCPRGWMHPTCGPCNCDVSGPBDCKNTSGECHKENHRRPBGSP 1920  
 DB 1861 YLGPYCESTRIDOCPRGWMHPTCGPCNCDVSGPBDCKNTSGECHKENHRRPBGSP 1920  
 QY 1921 CLICDCYPTGSLSRVCDPEDGQCPCKRGVIGROCDNCFPAVTTNGCEVNTDSCPRAT 1980  
 DB 1921 CLICDCYPTGSLSRVCDPEDGQCPCKRGVIGROCDNCFPAVTTNGCEVNTDSCPRAT 1980  
 QY 1981 BAGIWPRTFRGLPAAAPCPKGSFGTAVRHCDHRGMLPBNLFCNTSITFSELKGFABRL 2040  
 DB 1981 BAGIWPRTFRGLPAAAPCPKGSFGTAVRHCDHRGMLPBNLFCNTSITFSELKGFABRL 2040  
 QY 2041 QRBESGLDSGRSQQALLLNATOHNTAGVSGDVKAYAYOLATLHSHESQORFGLSAQ 2100  
 DB 2041 QRBESGLDSGRSQQALLLNATOHNTAGVSGDVKAYAYOLATLHSHESQORFGLSAQ 2100  
 QY 2101 DVAFTEMLRVGSALLDTANKRHWELIQTEGTAMLLQHYEAYASALQONMHTYLSPP 2160  
 DB 2101 DVAFTEMLRVGSALLDTANKRHWELIQTEGTAMLLQHYEAYASALQONMHTYLSPP 2160  
 QY 2161 TIYTPNIVISVRLDKNPAAGAKLPRYEALRGEPDPLETTVILPESVFEETPPVVRPAG 2220  
 DB 2161 TIYTPNIVISVRLDKNPAAGAKLPRYEALRGEPDPLETTVILPESVFEETPPVVRPAG 2220  
 QY 2221 PGEAROEPEELARORRHEBELSGEAVASVIYRTLGLLPHNYDPDRSLRVRKRPINT 2280  
 DB 2221 PGEAROEPEELARORRHEBELSGEAVASVIYRTLGLLPHNYDPDRSLRVRKRPINT 2280  
 QY 2281 PVSISVHDEBELPRALDKPVTVQFRLLETERKTPICVFMNHSILVSGTGMSARGCE 2340  
 DB 2281 PVSISVHDEBELPRALDKPVTVQFRLLETERKTPICVFMNHSILVSGTGMSARGCE 2340

QY 2341 VVFRNESHVSQCCNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFEPLTL 2400  
 DB 2341 VVFRNESHVSQCCNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFEPLTL 2400  
 QY 2401 IRIILSNQGRIRNLTALGLAQVFLGINADLPACTVIAIILHFLYCTESWALLE 2460  
 DB 2401 IRIILSNQGRIRNLTALGLAQVFLGINADLPACTVIAIILHFLYCTESWALLE 2460  
 QY 2461 ALHLRYALTEVADVNTGMRPFVYMLGMCVPAITGLAAGLDEBGVNPFCMLSTYDTLI 2520  
 DB 2461 ALHLRYALTEVADVNTGMRPFVYMLGMCVPAITGLAAGLDEBGVNPFCMLSTYDTLI 2520  
 QY 2521 WSPAGVPAVAVMSVFLYILARASCAAQROGFEKGPVSGLOPSFAVLLLSATWLLAL 2580  
 DB 2521 WSPAGVPAVAVMSVFLYILARASCAAQROGFEKGPVSGLOPSFAVLLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLPAFTNCIOGPPIFLSYVYLSKVRKALKLACGRKSPDALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLPAFTNCIOGPPIFLSYVYLSKVRKALKLACGRKSPDALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLYQPYGDSAGLSHSTRSGKSQPSYIPLRBSALNPGQPPGLG 2700  
 DB 2641 TSSYNCPSPYADGRLYQPYGDSAGLSHSTRSGKSQPSYIPLRBSALNPGQPPGLG 2700  
 QY 2701 DFGSLFLRQDQDHPDITDSDSLSDLDQSGSVASTHSDBSEEBEERBEBAFPGEG 2760  
 DB 2701 DFGSLFLRQDQDHPDITDSDSLSDLDQSGSVASTHSDBSEEBEERBEBAFPGEG 2760  
 QY 2761 WDSLIGPABRLPLHSTRDQGPGRKAPWPDRETTAKESGNAAPERLRENDALSR 2820  
 DB 2761 WDSLIGPABRLPLHSTRDQGPGRKAPWPDRETTAKESGNAAPERLRENDALSR 2820  
 QY 2821 EGSILAPLPGSSAQPHKGLIKKCLPTISEKSLPLPLEQCTGSSRGSASBSRGCGPP 2880  
 DB 2821 EGSILAPLPGSSAQPHKGLIKKCLPTISEKSLPLPLEQCTGSSRGSASBSRGCGPP 2880  
 QY 2881 RPPRQSIQEOQNGVPIAMSIKAGTVDESSGSEFFLPFNHL 2923  
 DB 2881 RPPRQSIQEOQNGVPIAMSIKAGTVDESSGSEFFLPFNHL 2923

RESULT 3  
 US-10-174-677-29  
 ; Sequence 29, Application US/10174677  
 ; Publication No. US20030190704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xie, Ting  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR  
 ; FILE REFERENCE: 40716(1P-012)  
 ; CURRENT APPLICATION NUMBER: US/10/174.677  
 ; CURRENT FILING DATE: 2002-06-19  
 ; NUMBER OF SEQ ID NOS: 117  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-174-677-29

Query Match 100.0%; Score 15545; DB 12; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPATGVPLPPTPEPLLLLLLLLPPLLDQVGPCHSLGSRGSSGACAPMGLCPSS 60  
 DB 1 MSPATGVPLPPTPEPLLLLLLLLPPLLDQVGPCHSLGSRGSSGACAPMGLCPSS 60  
 QY 61 SASNMLTYSRCKRDAGTBLTGHLVPHADGLRWYCESEBAHPLPAPBECGPMSCLLGIG 120  
 DB 61 SASNMLTYSRCKRDAGTBLTGHLVPHADGLRWYCESEBAHPLPAPBECGPMSCLLGIG 120  
 QY 121 GHLSPQSKLTPEBRPCLKAPRLRCOSGCLAQAPGLRAGERSPESSLGRRGRNNTAPO 180

Db 121 GHLSPQKLTLPBEPCLAPRLRCQSCKLAQAPGLRAGENSPEESLGRRKRNVTAPQ 180  
 Qy 181 FOPSPYQATVPENOPAGTPVVASLRAIDPDEGEAGLEXTMALFDRSNOFESLDPTVGA 240  
 Db 181 FOPSPYQATVPENOPAGTPVVASLRAIDPDEGEAGLEXTMALFDRSNOFESLDPTVGA 240  
 Qy 241 VTLAEELREKSTHVFVTAQDHGMPRRSALATITLVTNDNDPVPBEOETKESLRE 300  
 Db 241 VTLAEELREKSTHVFVTAQDHGMPRRSALATITLVTNDNDPVPBEOETKESLRE 300  
 Qy 301 NLEVEYEVLTARAATGDAPPNANILYRLLEGSGSPSEVEFIDPSSGVIRTRGPVDRREV 360  
 Db 301 NLEVEYEVLTARAATGDAPPNANILYRLLEGSGSPSEVEFIDPSSGVIRTRGPVDRREV 360  
 Qy 361 ESYQVLTASDQGRDPGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREBVTGPAPYL 420  
 Db 361 ESYQVLTASDQGRDPGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREBVTGPAPYL 420  
 Qy 421 VTAASDRDGSNAVHYSIMSGNARQFYLDAGTGMALDVVSPLDVETTKETLRVAADGG 480  
 Db 421 VTAASDRDGSNAVHYSIMSGNARQFYLDAGTGMALDVVSPLDVETTKETLRVAADGG 480  
 Qy 481 RPLPSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADAGDNRL 540  
 Db 481 RPLPSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADAGDNRL 540  
 Qy 541 EYRLAGVGHDPPTINNCGWISVAABLDREVDYFSGVEARDHGTALTAASVSVT 600  
 Db 541 EYRLAGVGHDPPTINNCGWISVAABLDREVDYFSGVEARDHGTALTAASVSVT 600  
 Qy 601 LDVNDNNTFPOPEVTVLNEBDAVAGTSVTVVSAVDRDAHSVITVQITSGTNRNFSITS 660  
 Db 601 LDVNDNNTFPOPEVTVLNEBDAVAGTSVTVVSAVDRDAHSVITVQITSGTNRNFSITS 660  
 Qy 661 OSGGGLVSLALPLDYKLEROYLAATASDGTRODPAQIVANVTANTHRPVSQSHVTYN 720  
 Db 661 OSGGGLVSLALPLDYKLEROYLAATASDGTRODPAQIVANVTANTHRPVSQSHVTYN 720  
 Qy 721 VNEBRPACTTVVLISATDEBTGENARITYFMEDSIPOFRIDADGAVTTOAELDYEDVS 780  
 Db 721 VNEBRPACTTVVLISATDEBTGENARITYFMEDSIPOFRIDADGAVTTOAELDYEDVS 780  
 Qy 781 YTLAITADNGIPKQSDPTTYLEIIVNDVNDNAPOFLRBSYOGSYVEDVPPTSVLQISAT 840  
 Db 781 YTLAITADNGIPKQSDPTTYLEIIVNDVNDNAPOFLRBSYOGSYVEDVPPTSVLQISAT 840  
 Qy 841 DRDSGLNRRVFTYFOGDDGDDGFIVESTSGIVRTLRDLRENVAAQVLRAYAVDKGMP 900  
 Db 841 DRDSGLNRRVFTYFOGDDGDDGFIVESTSGIVRTLRDLRENVAAQVLRAYAVDKGMP 900  
 Qy 901 ARTMEVTVTVLVDVNDNPVFEODEFDVFEENSPIGLAVARVATATDEBGTNAQIMYQI 960  
 Db 901 ARTMEVTVTVLVDVNDNPVFEODEFDVFEENSPIGLAVARVATATDEBGTNAQIMYQI 960  
 Qy 961 VEGNIPETVPODIPSGELTALVDLDYEDRPRVUVIQTASAPLYSRATVAVHRLDNRNP 1020  
 Db 961 VEGNIPETVPODIPSGELTALVDLDYEDRPRVUVIQTASAPLYSRATVAVHRLDNRNP 1020  
 Qy 1021 PVLGNFELLFNNYVTRSSSPFGAIGRVPADPDIPSLTYSFPRGNELSIVYLLNASTG 1080  
 Db 1021 PVLGNFELLFNNYVTRSSSPFGAIGRVPADPDIPSLTYSFPRGNELSIVYLLNASTG 1080  
 Qy 1081 ELKLSRALDNNRPLEAIVSVLSDGVHSVTAQCALRTVITIDEMLTHTSITLRLDMSPER 1140  
 Db 1081 ELKLSRALDNNRPLEAIVSVLSDGVHSVTAQCALRTVITIDEMLTHTSITLRLDMSPER 1140  
 Qy 1141 FLSPILGLFIOAVATAATATPPDHVVVENVORDTAPGHIINVSLSVQPPGPGGPPFL 1200  
 Db 1141 FLSPILGLFIOAVATAATATPPDHVVVENVORDTAPGHIINVSLSVQPPGPGGPPFL 1200  
 Qy 1201 PSEDLQERLYNRSILTAISAQVLPFDNICTLRBPCEMYRCSVLAEPDSAPFIASS 1260  
 Db 1201 PSEDLQERLYNRSILTAISAQVLPFDNICTLRBPCEMYRCSVLAEPDSAPFIASS 1260

Qy 1261 VLFRIHPVGLRCRCPGFFGADYCEITEVDLCYRPPCGPHRCRSREGYTCLCRDGYTG 1320  
 Db 1261 VLFRIHPVGLRCRCPGFFGADYCEITEVDLCYRPPCGPHRCRSREGYTCLCRDGYTG 1320  
 Qy 1321 EHCVSARSRCRTPGVCNKGCTCVNLLVGFKCDPESGDFEKPVCQVTTSPASFTTF 1380  
 Db 1321 EHCVSARSRCRTPGVCNKGCTCVNLLVGFKCDPESGDFEKPVCQVTTSPASFTTF 1380  
 Qy 1381 RGLRORHFTLALSFATKBRDGLLYNRFMEKHDPALEVIQOVOLTSAGSSTTVS 1440  
 Db 1381 RGLRORHFTLALSFATKBRDGLLYNRFMEKHDPALEVIQOVOLTSAGSSTTVS 1440  
 Qy 1441 PFVPGVSDGQMTVOLKXVYKPELLGQTLPOGSEOKVAVVTVDGCDTGAALRFSYLG 1500  
 Db 1441 PFVPGVSDGQMTVOLKXVYKPELLGQTLPOGSEOKVAVVTVDGCDTGAALRFSYLG 1500  
 Qy 1501 NYSCAAGTQGGSKSIDLTPILLGGVPDLPSFVPRMQFVGCNENLOVDSHIMAD 1560  
 Db 1501 NYSCAAGTQGGSKSIDLTPILLGGVPDLPSFVPRMQFVGCNENLOVDSHIMAD 1560  
 Qy 1561 FIANNGTVPGPCPAKQVCDNNTCHNGGTCVNOMDAPSCBCLPGRGKSCAQEMANPOHFL 1620  
 Db 1561 FIANNGTVPGPCPAKQVCDNNTCHNGGTCVNOMDAPSCBCLPGRGKSCAQEMANPOHFL 1620  
 Qy 1621 GSSLVANHGLSLPISQPMWLSLMEFTRQADGVLLQAITRGRSTITLQIRGHVMSVEGT 1680  
 Db 1621 GSSLVANHGLSLPISQPMWLSLMEFTRQADGVLLQAITRGRSTITLQIRGHVMSVEGT 1680  
 Qy 1681 GLOASSRLRPGRANGDMMHQAOLALGASGPGHAIISPTYGQORABGNTGPRHLGHL 1740  
 Db 1681 GLOASSRLRPGRANGDMMHQAOLALGASGPGHAIISPTYGQORABGNTGPRHLGHL 1740  
 Qy 1741 NITVGGIRPGAGVARGFRGCTLOGVRSDTPBGNLSIDPSHGSINVBQCSLPDPCDN 1800  
 Db 1741 NITVGGIRPGAGVARGFRGCTLOGVRSDTPBGNLSIDPSHGSINVBQCSLPDPCDN 1800  
 Qy 1801 PCPANSYCSNDMDSYSCDPGYGDMCTVCDLNPCEHOSVCTRKKSADHGVTCECPN 1860  
 Db 1801 PCPANSYCSNDMDSYSCDPGYGDMCTVCDLNPCEHOSVCTRKKSADHGVTCECPN 1860  
 Qy 1861 YLGPYCESTRIDPCPRMWMHPTCGPCNCDVSKGFPDCKTSGECHKRNHRPSPSP 1920  
 Db 1861 YLGPYCESTRIDPCPRMWMHPTCGPCNCDVSKGFPDCKTSGECHKRNHRPSPSP 1920  
 Qy 1921 CLLCDYPTGSLSRVCDPEBDGQCPCKRGVIGROCDRCNPFABVTTNGCEVNDSCRAI 1980  
 Db 1921 CLLCDYPTGSLSRVCDPEBDGQCPCKRGVIGROCDRCNPFABVTTNGCEVNDSCRAI 1980  
 Qy 1981 EAGIWMERTFGLPAAAPCEKSGFGTAVRHCDERHGLPBNLFNCTSIITSELSKGAERT 2040  
 Db 1981 EAGIWMERTFGLPAAAPCEKSGFGTAVRHCDERHGLPBNLFNCTSIITSELSKGAERT 2040  
 Qy 2041 QRNESGLDSRQOALLLNATQHTAGYGSVYKAYOULATLILAHESYQKRGFGLSATQ 2100  
 Db 2041 QRNESGLDSRQOALLLNATQHTAGYGSVYKAYOULATLILAHESYQKRGFGLSATQ 2100  
 Qy 2101 DVHFTBMLAVGSALDPTANKRMELIQQTEGTAMLOHYEAYASALQNMHHTYISPF 2160  
 Db 2101 DVHFTBMLAVGSALDPTANKRMELIQQTEGTAMLOHYEAYASALQNMHHTYISPF 2160  
 Qy 2161 TIVTPNIVISVRLDKGNFAGATLPRYBALRGOBPDLETTVILPESVRETPPVVAPAG 2220  
 Db 2161 TIVTPNIVISVRLDKGNFAGATLPRYBALRGOBPDLETTVILPESVRETPPVVAPAG 2220  
 Qy 2221 PGEAOEPEELARBORHPELSQGEAAVASVITYTLAGLPHNDPDKRSIRVPRPINT 2280  
 Db 2221 PGEAOEPEELARBORHPELSQGEAAVASVITYTLAGLPHNDPDKRSIRVPRPINT 2280  
 Qy 2281 PVSISVHDEBELPRLADKPVTVQFRLLETERTKPCVFMNHSILVSGTGMARSARGE 2340  
 Db 2281 PVSISVHDEBELPRLADKPVTVQFRLLETERTKPCVFMNHSILVSGTGMARSARGE 2340

```

QY 2341 VVFRNSHVSQCQNMNTSPAVIMDVSRRENSEILPLKTLTYVALGVTLAALLTFFELTL 2400
DB 2341 VVFRNSHVSQCQNMNTSPAVIMDVSRRENSEILPLKTLTYVALGVTLAALLTFFELTL 2400
QY 2401 LRLRNSHNGIRRNLTALGLAQVFLGIGNADLPACTVLAIIHLHYLCFPSALLL 2460
DB 2401 LRLRNSHNGIRRNLTALGLAQVFLGIGNADLPACTVLAIIHLHYLCFPSALLL 2460
QY 2461 ALHLRYALTEVRDVNTGPMRFYTMGMGVPAPITGLAVGLDPGYNPFCMLSIYDTLI 2520
DB 2461 ALHLRYALTEVRDVNTGPMRFYTMGMGVPAPITGLAVGLDPGYNPFCMLSIYDTLI 2520
QY 2521 WSPAGVAVAVSVKSVFLYIIAARASCAARQGFEGKGVSGIQLPSFAVILLLSATWLLAL 2580
DB 2521 WSPAGVAVAVSVKSVFLYIIAARASCAARQGFEGKGVSGIQLPSFAVILLLSATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFAICNCICQGFPIFLSVYVLSKVRKALKACSRKPSDDPALTTKSTL 2640
DB 2581 LSVNSDTLLFHYLFAICNCICQGFPIFLSVYVLSKVRKALKACSRKPSDDPALTTKSTL 2640
QY 2641 TSSYNCPSPYADRLYQPYDSAGSLHSTSRSGKSPSYIPLLRSEALNPQCGPGLG 2700
DB 2641 TSSYNCPSPYADRLYQPYDSAGSLHSTSRSGKSPSYIPLLRSEALNPQCGPGLG 2700
QY 2701 DPQSLFLBEGDDQHDPTDSDDLJLBDQSGSYASTHSDSEEEBEEBAAPFGEQ 2760
DB 2701 DPQSLFLBEGDDQHDPTDSDDLJLBDQSGSYASTHSDSEEEBEEBAAPFGEQ 2760
QY 2761 WBSLLPBGARLPLHSTPKXGPGPGKAPWPGDGTAKSSGNGAPBEBLRNGALR 2820
DB 2761 WBSLLPBGARLPLHSTPKXGPGPGKAPWPGDGTAKSSGNGAPBEBLRNGALR 2820
QY 2821 BGSIGLPGSSAOPHKGILKKKCLPTISEKSLRLPLBQCTSSSGSSASBGSRGAPP 2880
DB 2821 BGSIGLPGSSAOPHKGILKKKCLPTISEKSLRLPLBQCTSSSGSSASBGSRGAPP 2880
QY 2881 RPPRQSLQQLNGVMPIMSIKAGTYDEDSGSEPLFFNFLLH 2923
DB 2881 RPPRQSLQQLNGVMPIMSIKAGTYDEDSGSEPLFFNFLLH 2923

```

```

RESULT 4
US-10-120-801-53
; Sequence 53, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkete, Richard
; APPLICANT: Padigattu, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehrahan, Ruad
; APPLICANT: Topper, James N
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120, 801
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241

```

```

; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 53
; LENGTH: 2923
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-120-801-53

```

```

Query Match 100.0%; Score 15545; DB 12; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MRSRATGYPLPPPPPLMLMLMLLPPPLDQVQPCSLSGRSGSGACAPMGLCPSS 60
DB 1 MRSRATGYPLPPPPPLMLMLMLLPPPLDQVQPCSLSGRSGSGACAPMGLCPSS 60
QY 61 SASNMLYTSRCRDAGTELGHVPHDGLRVWCPESEAHPLPPAPGCPMGCGLGIG 120
DB 61 SASNMLYTSRCRDAGTELGHVPHDGLRVWCPESEAHPLPPAPGCPMGCGLGIG 120
QY 121 GHLSPOGKLTLPBEPHCILKAPRLRCQSCKLAQAPGLARGBSPBSLGGRRRNVTAPQ 180
DB 121 GHLSPOGKLTLPBEPHCILKAPRLRCQSCKLAQAPGLARGBSPBSLGGRRRNVTAPQ 180
QY 181 FPPPSYQATVPENQAGTPVASTRAIDPDEBAGLRTYTMALPBRNPFSLDPTGCA 240
DB 181 FPPPSYQATVPENQAGTPVASTRAIDPDEBAGLRTYTMALPBRNPFSLDPTGCA 240
QY 241 VTTAEELRETKSTVFVTAQDHPRRSALATLTITVTDNDHPVPEQGEYESLRE 300
DB 241 VTTAEELRETKSTVFVTAQDHPRRSALATLTITVTDNDHPVPEQGEYESLRE 300
QY 301 NLEVGVEVLTVATNDGAPPNANILYRLLEGSGSSPSEVFIDPBGVITRGPVDRBEV 360
DB 301 NLEVGVEVLTVATNDGAPPNANILYRLLEGSGSGSPSEVFIDPBGVITRGPVDRBEV 360
QY 361 ESYQULTVBSODGRPGPSTTAAPFLSVEDNDNAPQSEKRYVVOVEDTYPGAPVLR 420
DB 361 ESYQULTVBSODGRPGPSTTAAPFLSVEDNDNAPQSEKRYVVOVEDTYPGAPVLR 420
QY 421 VTASDRDKSNNAVHYSIMSGNARQFYLDAGTGLDVVSPLDVETTKETTLRVAAQDGG 480
DB 421 VTASDRDKSNNAVHYSIMSGNARQFYLDAGTGLDVVSPLDVETTKETTLRVAAQDGG 480
QY 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLESVPPLGYLVHQAIDADAGNARL 540
DB 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLESVPPLGYLVHQAIDADAGNARL 540
QY 541 EYRLAGVGHDPFTTNNNGWISVAEILDBREVDSYSGVBRDGTPLTASASVTV 600
DB 541 EYRLAGVGHDPFTTNNNGWISVAEILDBREVDSYSGVBRDGTPLTASASVTV 600
QY 601 LDVNDNNPFTQPEXTVRLNEDAAVGTSVTVSAVDRDAHSYITQITSGNTRNFSITS 660
DB 601 LDVNDNNPFTQPEXTVRLNEDAAVGTSVTVSAVDRDAHSYITQITSGNTRNFSITS 660
QY 661 QSGGGIVSLAPLDYKLERQYTLATASDGTRODTAQIVVNTDANTRPVQSSHYYN 720
DB 661 QSGGGIVSLAPLDYKLERQYTLATASDGTRODTAQIVVNTDANTRPVQSSHYYN 720
QY 721 VNEDRPAGTIVVLSATBEDTGENARITYFEMDSIPQIRIDAGTAVTTQAALDEDQVS 780
DB 721 VNEDRPAGTIVVLSATBEDTGENARITYFEMDSIPQIRIDAGTAVTTQAALDEDQVS 780
QY 781 YTLAITARDNGIPQSDTYYLEILVNDVNDNAPQFLRDSYQGSYVEDVPFTSVLQISAT 840
DB 781 YTLAITARDNGIPQSDTYYLEILVNDVNDNAPQFLRDSYQGSYVEDVPFTSVLQISAT 840

```

QY 841 DEDSGINGEFTTPOGGDDGDDGFIVESTSGIVTRLRRLREBNVAQVYLRAVAVDKGMP 900  
 DB 841 DEDSGINGEFTTPOGGDDGDDGFIVESTSGIVTRLRRLREBNVAQVYLRAVAVDKGMP 900  
 QY 901 ATTPMEVTVTVLDVNDNPFVEFDEDEPDVYEEKSPGLAVARVATYDDEBGTNAQIMYOI 960  
 DB 901 ATTPMEVTVTVLDVNDNPFVEFDEDEPDVYEEKSPGLAVARVATYDDEBGTNAQIMYOI 960  
 QY 961 VEGNIEVEVOLDI FSGELTALVDLDEDEPEYVLVQATSAPLVSRATYHVLBDNDP 1020  
 DB 961 VEGNIEVEVOLDI FSGELTALVDLDEDEPEYVLVQATSAPLVSRATYHVLBDNDP 1020  
 QY 1021 PYLGNFEILFNNVYTNRSSPPGAGIGRAYPANDPDISLTYSPERGNELSVLINAATG 1080  
 DB 1021 PYLGNFEILFNNVYTNRSSPPGAGIGRAYPANDPDISLTYSPERGNELSVLINAATG 1080  
 QY 1081 EKLSTRALDNNRPLKAINSVVSDGHSHTAOCALAVTITDBMLTHSITTLBDMSPER 1140  
 DB 1081 EKLSTRALDNNRPLKAINSVVSDGHSHTAOCALAVTITDBMLTHSITTLBDMSPER 1140  
 QY 1141 FLSPLGLFIQAVATLATPPDHVVVFNQORDTAPAGHILNVSLSVQPPGGGPPPL 1200  
 DB 1141 FLSPLGLFIQAVATLATPPDHVVVFNQORDTAPAGHILNVSLSVQPPGGGPPPL 1200  
 QY 1201 PSEDIQERLYNRSLSLTAISAQVLPEDNCLIREPCENYMRCSVLPDSSAPFIASS 1260  
 DB 1201 PSEDIQERLYNRSLSLTAISAQVLPEDNCLIREPCENYMRCSVLPDSSAPFIASS 1260  
 QY 1261 VLFRRPHVVGJRCRCRPFPTDGYCETEVDLCSRRCPHRCRBRGGYTCLCDDGYG 1320  
 DB 1261 VLFRRPHVVGJRCRCRPFPTDGYCETEVDLCSRRCPHRCRBRGGYTCLCDDGYG 1320  
 QY 1321 EHCESVARSRCPTGVCNKGCTCVNLLVGFKDCPSGDPEKPYCOVTRSPFAPSFTF 1380  
 DB 1321 EHCESVARSRCPTGVCNKGCTCVNLLVGFKDCPSGDPEKPYCOVTRSPFAPSFTF 1380  
 QY 1381 RGLRORFHTLTLSPATYKERDGLLYNGRFNEKHDFVALEVIQEOVLTFSGESTTVS 1440  
 DB 1381 RGLRORFHTLTLSPATYKERDGLLYNGRFNEKHDFVALEVIQEOVLTFSGESTTVS 1440  
 QY 1441 PVPVGVSGOMHTVQKXKYNKPLGOTGLPGSPBOKXAVVTVNGCDPVALRGSVYG 1500  
 DB 1441 PVPVGVSGOMHTVQKXKYNKPLGOTGLPGSPBOKXAVVTVNGCDPVALRGSVYG 1500  
 QY 1501 NYSCAOGTQGSKSLDLTGPULLGGVPLDESPFVRARQFVGCNRLQVDSRIHMDAD 1560  
 DB 1501 NYSCAOGTQGSKSLDLTGPULLGGVPLDESPFVRARQFVGCNRLQVDSRIHMDAD 1560  
 QY 1561 FIANNGTVPGPCAKKNVCSNTCHNGGTCVNOMDAFSCCEPLGFGKSCAQEMANPOHFL 1620  
 DB 1561 FIANNGTVPGPCAKKNVCSNTCHNGGTCVNOMDAFSCCEPLGFGKSCAQEMANPOHFL 1620  
 QY 1621 GSSLVAMHGLSLPIISOPWYLSLMPFTRROADGVLLAITGSRSTITLQAREGVMLSVBET 1680  
 DB 1621 GSSLVAMHGLSLPIISOPWYLSLMPFTRROADGVLLAITGSRSTITLQAREGVMLSVBET 1680  
 QY 1681 GLQASSLRLEPRANDGDWHHAQLALGASGPGHAILSPDYGOQBAEGNLGRLHGLHS 1740  
 DB 1681 GLQASSLRLEPRANDGDWHHAQLALGASGPGHAILSPDYGOQBAEGNLGRLHGLHS 1740  
 QY 1741 NITVGGIPGPAGVARGFPGCTQGVRSVTPREGVNSLDBSHESINVEQCSLPDPCDSN 1800  
 DB 1741 NITVGGIPGPAGVARGFPGCTQGVRSVTPREGVNSLDBSHESINVEQCSLPDPCDSN 1800  
 QY 1801 PCPANSYCSNDMSYSCSDPGYVGDNCNVCDLNPCEHOSVCTKPSAPHYTORCPNP 1860  
 DB 1801 PCPANSYCSNDMSYSCSDPGYVGDNCNVCDLNPCEHOSVCTKPSAPHYTORCPNP 1860  
 QY 1861 YLGAPYCETRIDQPCPRGMWGHPTGFCPCNDVSKGPDPCNDKTSGBCHKENHYRPPGSP 1920  
 DB 1861 YLGAPYCETRIDQPCPRGMWGHPTGFCPCNDVSKGPDPCNDKTSGBCHKENHYRPPGSP 1920  
 QY 1921 CLCOCYPTGSLSRVCDPEDGQPCPKPGVIGRQCDRCNPNPABVTTNGCEVNYDSCPRAI 1980

DB 1921 CLCOCYPTGSLSRVCDPEDGQPCPKPGVIGRQCDRCNPNPABVTTNGCEVNYDSCPRAI 1980  
 QY 1981 EAGIWWPRTFRGLPAAPCPKPGSGFCTAVRHCDEHGRMLPBNLFNCTSTPSELKPAERL 2040  
 DB 1981 EAGIWWPRTFRGLPAAPCPKPGSGFCTAVRHCDEHGRMLPBNLFNCTSTPSELKPAERL 2040  
 QY 2041 QNBSGLSGRSQOLALLRNATOTAGYPGSDVVAVALQATRLAHSTOGFGLSATQ 2100  
 DB 2041 QNBSGLSGRSQOLALLRNATOTAGYPGSDVVAVALQATRLAHSTOGFGLSATQ 2100  
 QY 2101 DVHFTENILRVGSALLDTANKRMELIOOTEGITAMLLQHEEYASALANNRHTYLSPF 2160  
 DB 2101 DVHFTENILRVGSALLDTANKRMELIOOTEGITAMLLQHEEYASALANNRHTYLSPF 2160  
 QY 2161 TIVTNTIVISVVRLDKGNFAKAPRYALNGBOPDLETTYILPESVRETTPVVRPAG 2220  
 DB 2161 TIVTNTIVISVVRLDKGNFAKAPRYALNGBOPDLETTYILPESVRETTPVVRPAG 2220  
 QY 2221 PGEAOPRELARORRHELSQGEAVASVIYRTLAGLLPHNYDPDKSLRVKXPIINT 2280  
 DB 2221 PGEAOPRELARORRHELSQGEAVASVIYRTLAGLLPHNYDPDKSLRVKXPIINT 2280  
 QY 2281 PVSISVHDEBELLPALDKPVTYQFLLJTEBRTKPICVFMNHSILVSGTGMGARGCE 2340  
 DB 2281 PVSISVHDEBELLPALDKPVTYQFLLJTEBRTKPICVFMNHSILVSGTGMGARGCE 2340  
 QY 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGELLPLKLTLYVALGVTLAALLTFEFTL 2400  
 DB 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGELLPLKLTLYVALGVTLAALLTFEFTL 2400  
 QY 2401 LRLRNSNCHIRRLTALGLAOLVFLGIGNOQDLPAFCVYIATILHFLVCTESWALLE 2460  
 DB 2401 LRLRNSNCHIRRLTALGLAOLVFLGIGNOQDLPAFCVYIATILHFLVCTESWALLE 2460  
 QY 2461 ALHLYRALTEYRDVNTGEMRFTYMLGWCVPAPITGLAVGLDPEGYNDPCMLSTYDTLI 2520  
 DB 2461 ALHLYRALTEYRDVNTGEMRFTYMLGWCVPAPITGLAVGLDPEGYNDPCMLSTYDTLI 2520  
 QY 2521 WSPFAPVAFVAMSVEPLYTLAARASCAAROGFPEKGVSGLOSPFALLLSATWLLAL 2580  
 DB 2521 WSPFAPVAFVAMSVEPLYTLAARASCAAROGFPEKGVSGLOSPFALLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLSTYVLSKVRKALKLACSRKSPDPALITKSTL 2640  
 DB 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLSTYVLSKVRKALKLACSRKSPDPALITKSTL 2640  
 QY 2641 TTSYNCPSPYADGRLYOPYGDSAGSLHSTSRGKSOPSYIPPLREBSALNPGQPPGLG 2700  
 DB 2641 TTSYNCPSPYADGRLYOPYGDSAGSLHSTSRGKSOPSYIPPLREBSALNPGQPPGLG 2700  
 QY 2701 DPGSLFLEGDOQOHPDPTDSDLSLEDDSGSYASTSSBSEBSEBSEBAAAPGBOG 2760  
 DB 2701 DPGSLFLEGDOQOHPDPTDSDLSLEDDSGSYASTSSBSEBSEBSEBAAAPGBOG 2760  
 QY 2761 WDSLILGPABRLPLHSTPKOGPGPGKAPMPGADGCTTAKSSGNGCAPERLRENGDALSR 2820  
 DB 2761 WDSLILGPABRLPLHSTPKOGPGPGKAPMPGADGCTTAKSSGNGCAPERLRENGDALSR 2820  
 QY 2821 EGSIGPLPGSSAOPHKGLKKKCLPTISBKSLRLPLBOCTGSSRGSSASBGRGPP 2880  
 DB 2821 EGSIGPLPGSSAOPHKGLKKKCLPTISBKSLRLPLBOCTGSSRGSSASBGRGPP 2880  
 QY 2881 RPPPROSLQEOQLNGVMPFAMSIXKATYDDEBSSGSEPLFFNPLH 2923  
 DB 2881 RPPPROSLQEOQLNGVMPFAMSIXKATYDDEBSSGSEPLFFNPLH 2923

RESULT 5  
 US-10-292-798-932  
 ; Sequence 932, Application us/10292798  
 ; Publication No. US20030235833A1  
 ; GENERAL INFORMATION:

/ APPLICANT: SUMA, MAKIKO  
 / APPLICANT: ASAI, KIYOSHI  
 / APPLICANT: AKIYAMA, YUTAKA  
 / APPLICANT: AUBERANT, HIROYUKI  
 / TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
 / FILE REFERENCE: 084335/166  
 / CURRENT APPLICATION NUMBER: US/10/292,798  
 / PRIOR FILING DATE: 2002-11-13  
 / PRIOR APPLICATION NUMBER: 10/017,161  
 / PRIOR FILING DATE: 2001-12-18  
 / PRIOR APPLICATION NUMBER: JP 2001-246789  
 / NUMBER OF SEQ ID NOS: 2070  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO: 932  
 / LENGTH: 2923  
 / TYPE: PRN  
 / ORGANISM: Homo sapiens  
 / US-10-292-798-932

Query Match 100.0%; Score 15545; DB 12; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSPTATGVP...  
 DB 1 MRSPTATGVP...  
 QY 61 SASNLWLYSR...  
 DB 61 SASNLWLYSR...  
 QY 121 GHLSPOGKLT...  
 DB 121 GHLSPOGKLT...  
 QY 181 POPPSQATV...  
 DB 181 POPPSQATV...  
 QY 241 VTTAAEELDR...  
 DB 241 VTTAAEELDR...  
 QY 301 NLEVGVEVL...  
 DB 301 NLEVGVEVL...  
 QY 361 ESTQVTV...  
 DB 361 ESTQVTV...  
 QY 421 VTSADBDK...  
 DB 421 VTSADBDK...  
 QY 481 RPLSNVSG...  
 DB 481 RPLSNVSG...  
 QY 541 EYRLAGV...  
 DB 541 EYRLAGV...  
 QY 601 LDVNDNN...  
 DB 601 LDVNDNN...  
 QY 661 QSGGGV...  
 DB 661 QSGGGV...  
 QY 721 VNDRPAG...  
 DB 721 VNDRPAG...

DB 721 VNDRPAG...  
 QY 781 YTLATTA...  
 DB 781 YTLATTA...  
 QY 841 DRDSGLN...  
 DB 841 DRDSGLN...  
 QY 901 ARTPEV...  
 DB 901 ARTPEV...  
 QY 961 VEGNIPE...  
 DB 961 VEGNIPE...  
 QY 1021 PVLGNFE...  
 DB 1021 PVLGNFE...  
 QY 1081 EKLKSR...  
 DB 1081 EKLKSR...  
 QY 1141 FLSPILG...  
 DB 1141 FLSPILG...  
 QY 1201 PSEDLO...  
 DB 1201 PSEDLO...  
 QY 1261 VLFRIH...  
 DB 1261 VLFRIH...  
 QY 1321 EHCESV...  
 DB 1321 EHCESV...  
 QY 1381 RGLRQF...  
 DB 1381 RGLRQF...  
 QY 1441 PRVPGV...  
 DB 1441 PRVPGV...  
 QY 1501 NYS...  
 DB 1501 NYS...  
 QY 1561 FIANNG...  
 DB 1561 FIANNG...  
 QY 1621 GSSVAM...  
 DB 1621 GSSVAM...  
 QY 1681 GLQAS...  
 DB 1681 GLQAS...  
 QY 1741 NITVGG...  
 DB 1741 NITVGG...  
 QY 1801 PCPAN...  
 DB 1801 PCPAN...

1801 PCPANSYCSNDMDVSQCDPRTYGDNCINVCIDLANPCHQSVCTKRPKSAFHGTYCECPN 1860  
1861 YLGPYCEIRIDPCPCRGWGHPTGPCNCDVSKGPDPCDKTSGBCXKHNHYRPGSPPT 1920  
1861 YLGPYCEIRIDPCPCRGWGHPTGPCNCDVSKGPDPCDKTSGBCXKHNHYRPGSPPT 1920  
1921 CILCCYCPFGSLSRVDPEDGQCPCKPQYIGRCQCRCDNPPAEVTTNGCEVNYDSCPAI 1980  
1921 CILCCYCPFGSLSRVDPEDGQCPCKPQYIGRCQCRCDNPPAEVTTNGCEVNYDSCPAI 1980  
1921 CILCCYCPFGSLSRVDPEDGQCPCKPQYIGRCQCRCDNPPAEVTTNGCEVNYDSCPAI 1980  
1981 EAGIWMFRTFGLPAAAPCPKSGFGTAVRHCEHGMWLPNPFNCSTIFSEILKGAERL 2040  
1981 EAGIWMFRTFGLPAAAPCPKSGFGTAVRHCEHGMWLPNPFNCSTIFSEILKGAERL 2040  
2041 QNRSGLDGRSQOALALLRNATQHTAGYFGSDVKAVALRLLAHSTORGFGLSATQ 2100  
2041 QNRSGLDGRSQOALALLRNATQHTAGYFGSDVKAVALRLLAHSTORGFGLSATQ 2100  
2101 DVHFTENILRVSSALLDTANKHMLIQOTEGTAMLOHTAYASALAQNMRHTYLSPP 2160  
2101 DVHFTENILRVSSALLDTANKHMLIQOTEGTAMLOHTAYASALAQNMRHTYLSPP 2160  
2161 TIVTENVIVSVRLDKGNFAGAKLPRYEALRGEOPDLETTVILPESVRETPPVVRPAG 2220  
2161 TIVTENVIVSVRLDKGNFAGAKLPRYEALRGEOPDLETTVILPESVRETPPVVRPAG 2220  
2221 PGEAOEPEELARORRHPELSGEAVASYIYRTLAGLPHNYDDPKSLRYPKPIINT 2280  
2221 PGEAOEPEELARORRHPELSGEAVASYIYRTLAGLPHNYDDPKSLRYPKPIINT 2280  
2281 PVSISVHDEDELLPRALDKPTVQFRLLETERKTPICVFNHNSILVSGTGMSARCE 2340  
2281 PVSISVHDEDELLPRALDKPTVQFRLLETERKTPICVFNHNSILVSGTGMSARCE 2340  
2341 VVFRNESHVSQCCNMHTSPAVLMDVSRRENGELLPKLTLYVALGTYLALLLTFPPLTL 2400  
2341 VVFRNESHVSQCCNMHTSPAVLMDVSRRENGELLPKLTLYVALGTYLALLLTFPPLTL 2400  
2401 LRILSNHOGIRRNLTAAIGLAQVFLGIGNADLPACTVTAIILHFLYCTPFWALL 2460  
2401 LRILSNHOGIRRNLTAAIGLAQVFLGIGNADLPACTVTAIILHFLYCTPFWALL 2460  
2461 ALHLVRLTEVRDVMNGPFRFYMMGMGVPATITGLAVGLDPBGVGNDFCWLSTYDTLI 2520  
2461 ALHLVRLTEVRDVMNGPFRFYMMGMGVPATITGLAVGLDPBGVGNDFCWLSTYDTLI 2520  
2521 MSFAPVAPAVMSYFLYLAAARASCAQROGFEEKGPGVSGLOPFAVILLISATWIAL 2580  
2521 MSFAPVAPAVMSYFLYLAAARASCAQROGFEEKGPGVSGLOPFAVILLISATWIAL 2580  
2581 LSVNSDTLLFHYLFATCNCIQGPFIFLSVYVLSKEVRKALKACSRKSPDPALTTKSTL 2640  
2581 LSVNSDTLLFHYLFATCNCIQGPFIFLSVYVLSKEVRKALKACSRKSPDPALTTKSTL 2640  
2641 TSSVNCBSPVADGRLYOPYGBSAGLSHSTRSGKQPSYIPLLEBBSALNPGQPPGIG 2700  
2641 TSSVNCBSPVADGRLYOPYGBSAGLSHSTRSGKQPSYIPLLEBBSALNPGQPPGIG 2700  
2701 DPGSLFLEGQDQHPDITDSDSLGLBEDQSGSYASTSSDSEBEEBEEBEEBAPGBOG 2760  
2701 DPGSLFLEGQDQHPDITDSDSLGLBEDQSGSYASTSSDSEBEEBEEBEEBAPGBOG 2760  
2761 WDSLILGPAERLPLHSTPDGPGPGKAPWPGDFGTAKESGNGAPRELRRENDALSR 2820  
2761 WDSLILGPAERLPLHSTPDGPGPGKAPWPGDFGTAKESGNGAPRELRRENDALSR 2820  
2821 EGSIGLPLGSSAOPBKGLIKKCLPTISEKSLALPLBQCTGSRGSSASBGSNGGPP 2880  
2821 EGSIGLPLGSSAOPBKGLIKKCLPTISEKSLALPLBQCTGSRGSSASBGSNGGPP 2880  
2881 RPPRQSLQEBQLNGVPIAMSIKAGTVDEDSGSEFLFNFPLH 2923  
2881 RPPRQSLQEBQLNGVPIAMSIKAGTVDEDSGSEFLFNFPLH 2923  
2881 RPPRQSLQEBQLNGVPIAMSIKAGTVDEDSGSEFLFNFPLH 2923

RESULT 6  
US-10-225-567A-524  
Sequence 524, Application US/10225567A  
Publication No. US20030113798A1  
GENERAL INFORMATION:  
APPLICANT: Lifespan Biosciences  
APPLICANT: Brown, Joseph P.  
APPLICANT: Burner, Glenna C.  
APPLICANT: Roush, Christine L.  
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 1920-4-4  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/257,144  
PRIOR FILING DATE: 2000-12-19  
NUMBER OF SEQ ID NOS: 2292  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 524  
LENGTH: 2923  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-225-567A-524  
Query Match 100.0%; Score 15545; DB 15; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MRSPTATGVPPLTPPPPLLLLLLLPPPLLDQVGPCKSLGSRGSSGACAPMGWLCPS 60  
1 MRSPTATGVPPLTPPPPLLLLLLLPPPLLDQVGPCKSLGSRGSSGACAPMGWLCPS 60  
61 SASNMLYTSRCRAGTELGHVPHHDGLVWCPEBAHPLPAPGCPMSRLIGIG 120  
61 SASNMLYTSRCRAGTELGHVPHHDGLVWCPEBAHPLPAPGCPMSRLIGIG 120  
121 GHLSPOGKLTLPBEPCLKAPRLRCOSCKLAQAAGLRAGRSPESISGRKRKRVNTAPQ 180  
121 GHLSPOGKLTLPBEPCLKAPRLRCOSCKLAQAAGLRAGRSPESISGRKRKRVNTAPQ 180  
181 FQPSYQATVPENOPAGTPVASLALIDPDGEGARLEYMDALDSSNOFFSLDPVTGA 240  
181 FQPSYQATVPENOPAGTPVASLALIDPDGEGARLEYMDALDSSNOFFSLDPVTGA 240  
181 FQPSYQATVPENOPAGTPVASLALIDPDGEGARLEYMDALDSSNOFFSLDPVTGA 240  
241 VTTAEELDRKETSHTVRVTAADHGMPSALATLTLYMDTNDHDPVBOQEKSLRE 300  
241 VTTAEELDRKETSHTVRVTAADHGMPSALATLTLYMDTNDHDPVBOQEKSLRE 300  
301 NLEVGVEVLTVRATDGDAPPANILYRLBSGSGSPSEVEIDPRSGVIRTRGVDBEEV 360  
301 NLEVGVEVLTVRATDGDAPPANILYRLBSGSGSPSEVEIDPRSGVIRTRGVDBEEV 360  
361 ESYQLTVEASDQGRDPPRSTTAAVLSVEDDNDNAPQFSEKRYVVOVRBDVTFGAYLR 420  
361 ESYQLTVEASDQGRDPPRSTTAAVLSVEDDNDNAPQFSEKRYVVOVRBDVTFGAYLR 420  
421 VTASDRKGSNAVHYHISMGNAAGQFYLDAGALNVSPLDVETTKETTLRYRADOOG 480  
421 VTASDRKGSNAVHYHISMGNAAGQFYLDAGALNVSPLDVETTKETTLRYRADOOG 480  
481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLSVPLGYLVLAHQALDADAGNARL 540  
481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLSVPLGYLVLAHQALDADAGNARL 540  
541 EYRLAGVGHDPPTTNGTGMISVAALDREEDVDFYFGYBAADHGPATLASASVTV 600  
541 EYRLAGVGHDPPTTNGTGMISVAALDREEDVDFYFGYBAADHGPATLASASVTV 600  
601 LDVVDNPNFTQEBEYVRLNEDAAVGSVTVSAVNDASVITYQTSGNTRRPSITS 660  
601 LDVVDNPNFTQEBEYVRLNEDAAVGSVTVSAVNDASVITYQTSGNTRRPSITS 660  
601 LDVVDNPNFTQEBEYVRLNEDAAVGSVTVSAVNDASVITYQTSGNTRRPSITS 660



```

QY 661 QSGGGLVSLALPDYTLERQYLAATASDCTRODTQIYVNTVDANTHAPVFOSSHATTN 720
Db 661 QSGGGLVSLALPDYTLERQYLAATASDCTRODTQIYVNTVDANTHAPVFOSSHATTN 720
QY 721 VNEBRPAGTTVVLISATDEDTGENARITYEMEDSIFQFRIDADTGA VTTQALDYEDQVS 780
Db 721 VNEBRPAGTTVVLISATDEDTGENARITYEMEDSIFQFRIDADTGA VTTQALDYEDQVS 780
QY 781 YTLAIFARDNGIPQKSDTTYLAILVNDVNDNAPQPLRDSYQSGSVYEDVPEFTSVLQISAT 840
Db 781 YTLAIFARDNGIPQKSDTTYLAILVNDVNDNAPQPLRDSYQSGSVYEDVPEFTSVLQISAT 840
QY 841 DRDSGLANGRYFTYFQGGDDGDDPFIYESISGIVRTLRDLRENVAAQYLAAYVNDGMP 900
Db 841 DRDSGLANGRYFTYFQGGDDGDDPFIYESISGIVRTLRDLRENVAAQYLAAYVNDGMP 900
QY 901 ARTPEMTVTVLVDVNNPPIVFEODEPFIYEENSPIGLAVARTATDPBGTNAQIMYQI 960
Db 901 ARTPEMTVTVLVDVNNPPIVFEODEPFIYEENSPIGLAVARTATDPBGTNAQIMYQI 960
QY 961 VEGNIDPEVFOJDI FSGELTALVDLYEDREBYVVIQATISAPLVSRATYHVLDDNDP 1020
Db 961 VEGNIDPEVFOJDI FSGELTALVDLYEDREBYVVIQATISAPLVSRATYHVLDDNDP 1020
QY 1021 PVLGNPELLENNVYTRSSSPFGGATGRVPAHPDISDSTYSFERGNEISLYLNASTG 1080
Db 1021 PVLGNPELLENNVYTRSSSPFGGATGRVPAHPDISDSTYSFERGNEISLYLNASTG 1080
QY 1081 ELKLSALDNNRPLEAIMSVLSDGVHSTPAOCALRVITITDEMTHSITLRLEDSMPER 1140
Db 1081 ELKLSALDNNRPLEAIMSVLSDGVHSTPAOCALRVITITDEMTHSITLRLEDSMPER 1140
QY 1141 FLSPLGLPLIOAVAATLATPPDHVVFNVRQDTDAPEGHILNYSLSVGOPGPGGPPPL 1200
Db 1141 FLSPLGLPLIOAVAATLATPPDHVVFNVRQDTDAPEGHILNYSLSVGOPGPGGPPPL 1200
QY 1201 PSSDLOERLYLANSLLTASAOVLPPDNI CLRPECNMRCVSVLRPSSAPFLASSS 1260
Db 1201 PSSDLOERLYLANSLLTASAOVLPPDNI CLRPECNMRCVSVLRPSSAPFLASSS 1260
QY 1261 VLFRPIHPVGLRCRCPGFTGDCETEVDLCYSRCPGPHGRSREGVTCACRDGYTG 1320
Db 1261 VLFRPIHPVGLRCRCPGFTGDCETEVDLCYSRCPGPHGRSREGVTCACRDGYTG 1320
QY 1321 EHCEVSARSGRCTPGVCNKGCTCNILVSGFKDCPSGDFEKPVCQVTTSPFAHSFITP 1380
Db 1321 EHCEVSARSGRCTPGVCNKGCTCNILVSGFKDCPSGDFEKPVCQVTTSPFAHSFITP 1380
QY 1381 RGLRORPHFTLALSPATKERDGLLYNGRFNEKHPFALEVIOEVOQLTFSSAGESTTVS 1440
Db 1381 RGLRORPHFTLALSPATKERDGLLYNGRFNEKHPFALEVIOEVOQLTFSSAGESTTVS 1440
QY 1441 PVPVGVSDQOMHTVOLKYTNKPLCQTGLPOGSEBOKVAVVVDGCDTGVALRFGSVLG 1500
Db 1441 PVPVGVSDQOMHTVOLKYTNKPLCQTGLPOGSEBOKVAVVVDGCDTGVALRFGSVLG 1500
QY 1501 NYSCAAQGTQSGSKSIDLTGPIILGGVPLPSSFPVYRMQFTVGCNKNLQVDSRHIDMAD 1560
Db 1501 NYSCAAQGTQSGSKSIDLTGPIILGGVPLPSSFPVYRMQFTVGCNKNLQVDSRHIDMAD 1560
QY 1561 FIANNNGVPGCPAKKVCNDNTNNGGTCYMONDAFSCBCLPFGGKSCAOEAMNPOHPL 1620
Db 1561 FIANNNGVPGCPAKKVCNDNTNNGGTCYMONDAFSCBCLPFGGKSCAOEAMNPOHPL 1620
QY 1621 GSSIVAHNGSLPLISQWYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVLSEGT 1680
Db 1621 GSSIVAHNGSLPLISQWYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVLSEGT 1680
QY 1681 GLQASSRLRPPGRANDQDHHQAOLALGASGPGHAIISFPYGOORABAGNIGPRLHGLHS 1740
Db 1681 GLQASSRLRPPGRANDQDHHQAOLALGASGPGHAIISFPYGOORABAGNIGPRLHGLHS 1740
QY 1741 NITVGGIPGAGVARGFRGLQGVRSVDTPEGVNSLDPHGESINVEQCSLPDPDCSN 1800
Db 1741 NITVGGIPGAGVARGFRGLQGVRSVDTPEGVNSLDPHGESINVEQCSLPDPDCSN 1800

Db 1741 NITVGGIPGAGVARGFRGLQGVRSVDTPEGVNSLDPHGESINVEQCSLPDPDCSN 1800
QY 1801 PCPNASYCNSMNDVSQSCDPGYDNCNTNVCNLPCEHOSVCTRKPAFHGYCECPN 1860
Db 1801 PCPNASYCNSMNDVSQSCDPGYDNCNTNVCNLPCEHOSVCTRKPAFHGYCECPN 1860
QY 1861 YLGPYCETRIDQPCRGWNGHPTCGPCNCDVSKGFPDPCNKTSGECHKENHYPGSPPT 1920
Db 1861 YLGPYCETRIDQPCRGWNGHPTCGPCNCDVSKGFPDPCNKTSGECHKENHYPGSPPT 1920
QY 1921 CLICCCYTRGSLSRVCDPEDQCPCKPGVITGRQCDRCNPNPAEYVTTNGCVNVYDSCPAI 1980
Db 1921 CLICCCYTRGSLSRVCDPEDQCPCKPGVITGRQCDRCNPNPAEYVTTNGCVNVYDSCPAI 1980
QY 1981 EAGIWMPPTRFGLPAAAPCPKPSFGTAARHCDERHGMPLPMLFNCTSIYFSELKFAERL 2040
Db 1981 EAGIWMPPTRFGLPAAAPCPKPSFGTAARHCDERHGMPLPMLFNCTSIYFSELKFAERL 2040
QY 2041 QNRESGLDSGSOQLALLIRNATOHTAGYFGSDVAVAYQALRLLAHESSTORGFLSATQ 2100
Db 2041 QNRESGLDSGSOQLALLIRNATOHTAGYFGSDVAVAYQALRLLAHESSTORGFLSATQ 2100
QY 2101 DVHFTENLLRVGSALLDTPANKHWEILIOQTSBGTAMLLQHYEAVASALANMRHTYLSPF 2160
Db 2101 DVHFTENLLRVGSALLDTPANKHWEILIOQTSBGTAMLLQHYEAVASALANMRHTYLSPF 2160
QY 2161 TIVTENVIVSVRLDKGNFAGAKLPRYALNGEQPDELTTVILPESYFRETPPVVRPAG 2220
Db 2161 TIVTENVIVSVRLDKGNFAGAKLPRYALNGEQPDELTTVILPESYFRETPPVVRPAG 2220
QY 2221 PGEAOPEBELARORRHPELSQGBAVASVITYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
Db 2221 PGEAOPEBELARORRHPELSQGBAVASVITYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
QY 2281 PVVVISVHDEBELRALDKPVTVOFRLLEBERTKPICVFNHNSIIVSGTGSGARGCE 2340
Db 2281 PVVVISVHDEBELRALDKPVTVOFRLLEBERTKPICVFNHNSIIVSGTGSGARGCE 2340
QY 2341 VVFRNESHVSQCNMHTSPAVALMDVSRRENGEILPLKTLTYVALGVTLAALLTFEFLTL 2400
Db 2341 VVFRNESHVSQCNMHTSPAVALMDVSRRENGEILPLKTLTYVALGVTLAALLTFEFLTL 2400
QY 2401 IALISNQHGRNRTALAGLAQVFLGINOADI PRACVIAIILHPLYLCTBEMALLE 2460
Db 2401 IALISNQHGRNRTALAGLAQVFLGINOADI PRACVIAIILHPLYLCTBEMALLE 2460
QY 2461 ALHLRYALTEVADVNTGPMRFYMLGMGPAPITGILAVGLDBEGXGNDPFCMLSIYDTLI 2520
Db 2461 ALHLRYALTEVADVNTGPMRFYMLGMGPAPITGILAVGLDBEGXGNDPFCMLSIYDTLI 2520
QY 2521 WSPFAPVAPAVSMVFLYTLAARASCAAORQGEKKGPVSGLOPSPAVILLISATWLLAL 2580
Db 2521 WSPFAPVAPAVSMVFLYTLAARASCAAORQGEKKGPVSGLOPSPAVILLISATWLLAL 2580
QY 2581 LSVNSDTLLFHTLFATKNCIOQPFITLSYVYLSKEVRLAKLACSRKSPDRLATTKSTL 2640
Db 2581 LSVNSDTLLFHTLFATKNCIOQPFITLSYVYLSKEVRLAKLACSRKSPDRLATTKSTL 2640
QY 2641 TSSYNCPSPYADGRLYQPGDSAGSLHSTRSGKQPSYIPTLAREBSALNPGQPPGIG 2700
Db 2641 TSSYNCPSPYADGRLYQPGDSAGSLHSTRSGKQPSYIPTLAREBSALNPGQPPGIG 2700
QY 2701 DPGSLFLBQDQDHPDSDSLSLBDDQSGSVASTHSDBEBEBEBEBEAPFPGGQ 2760
Db 2701 DPGSLFLBQDQDHPDSDSLSLBDDQSGSVASTHSDBEBEBEBEBEAPFPGGQ 2760
QY 2761 WDSILGPGAEPLPLHSTPDGGRGPGKAPWPEPDPCITTKESGNGAPREERLENDALSR 2820
Db 2761 WDSILGPGAEPLPLHSTPDGGRGPGKAPWPEPDPCITTKESGNGAPREERLENDALSR 2820
QY 2821 EGSIGPLPSSAQPHKGLIKKKCLPTISBKSILRLPLEQCTGSSRGSASBGRGPP 2880
Db 2821 EGSIGPLPSSAQPHKGLIKKKCLPTISBKSILRLPLEQCTGSSRGSASBGRGPP 2880

```



```

QY 1741 NITVGGIRPGAGVARGFRGLQGVRSVDPTEGVNSLDPHSGESINVEQCSLPDPCDSN 1800
Db 1741 NITVGGIRPGAGVARGFRGLQGVRSVDPTEGVNSLDPHSGESINVEQCSLPDPCDSN 1800
QY 1801 PCPANSICSDNDWMSYSCSCDPGYTGDNCTVCDLNPCEHOSVCTKRSAPHGTYTCCPEPN 1860
Db 1801 PCPANSICSDNDWMSYSCSCDPGYTGDNCTVCDLNPCEHOSVCTKRSAPHGTYTCCPEPN 1860
QY 1861 YLGPYCTRIDQPCPRGMWGHPTGCPNCDVSKGFPDDCNKTSGECHCKENHTRPPSPPT 1920
Db 1861 YLGPYCTRIDQPCPRGMWGHPTGCPNCDVSKGFPDDCNKTSGECHCKENHTRPPSPPT 1920
QY 1921 CLLCDCTPTGSLASRCDPEDQCPCKEGLGRQCDRCNPNFAEYTTNGCEVNTYDSCPRAI 1980
Db 1921 CLLCDCTPTGSLASRCDPEDQCPCKEGLGRQCDRCNPNFAEYTTNGCEVNTYDSCPRAI 1980
QY 1981 EAGIWMWRTFRGLPAAAPCPKRGSGTAVRHCDERHGMPLPMLFNCTISITSEBKGFARL 2040
Db 1981 EAGIWMWRTFRGLPAAAPCPKRGSGTAVRHCDERHGMPLPMLFNCTISITSEBKGFARL 2040
QY 2041 QRNBSGLDSGRSQQALILRNATQHTAGYFGSDVKVAYOLATRLAHSTORGFGLSATQ 2100
Db 2041 QRNBSGLDSGRSQQALILRNATQHTAGYFGSDVKVAYOLATRLAHSTORGFGLSATQ 2100
QY 2101 DVHFTENLAVGSALDTPANKRHMLIQOTEGETAMLQHYEAVASALQNMHTYLSPT 2160
Db 2101 DVHFTENLAVGSALDTPANKRHMLIQOTEGETAMLQHYEAVASALQNMHTYLSPT 2160
QY 2161 TIYTPNIVISVNLDKNPFAGAKLPRYBARGBPPLLETTVILPESEVFETPPVVRPAG 2220
Db 2161 TIYTPNIVISVNLDKNPFAGAKLPRYBARGBPPLLETTVILPESEVFETPPVVRPAG 2220
QY 2221 PGBAQEBELARQRRPELSQGEAVASVILYRTAGLPHANTDPDGRSLRVPRPIINT 2280
Db 2221 PGBAQEBELARQRRPELSQGEAVASVILYRTAGLPHANTDPDGRSLRVPRPIINT 2280
QY 2281 PVSISVHDEBELPRALDKPVTVQFLLTEERTKPICVFNMHSILVSGYGSARGCE 2340
Db 2281 PVSISVHDEBELPRALDKPVTVQFLLTEERTKPICVFNMHSILVSGYGSARGCE 2340
QY 2341 VVRNESHVSQCCNHNMTSPAVLMDVSRREGEILPLTLTVYVAGTVALALLTTPPLTL 2400
Db 2341 VVRNESHVSQCCNHNMTSPAVLMDVSRREGEILPLTLTVYVAGTVALALLTTPPLTL 2400
QY 2401 LRILRSNQHIRRNLTALGLAOLVFLGLINQADLPACTVIALHFLYLCTFSWALLE 2460
Db 2401 LRILRSNQHIRRNLTALGLAOLVFLGLINQADLPACTVIALHFLYLCTFSWALLE 2460
QY 2461 ALHLRYALTEVRDVTGPMRFYTMGLGVPATITGLAVGLDPGYNPDFCWLSTYDTLI 2520
Db 2461 ALHLRYALTEVRDVTGPMRFYTMGLGVPATITGLAVGLDPGYNPDFCWLSTYDTLI 2520
QY 2521 WSPRAGPAPFVMSVFLYILAAASCAARQGEKGVSGLOPSPFVILLISATWLLAL 2580
Db 2521 WSPRAGPAPFVMSVFLYILAAASCAARQGEKGVSGLOPSPFVILLISATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFATCNCTIGPFIPLSYVVLSEKRYAKLKLACSRKSPDPAITTKSTL 2640
Db 2581 LSVNSDTLLFHYLFATCNCTIGPFIPLSYVVLSEKRYAKLKLACSRKSPDPAITTKSTL 2640
QY 2641 TTSYNCPSPYADGRLVOPYGDSAGLSHSTRSGKSPSYPLILREBSALNPQGGPPGLG 2700
Db 2641 TTSYNCPSPYADGRLVOPYGDSAGLSHSTRSGKSPSYPLILREBSALNPQGGPPGLG 2700
QY 2701 DPGSLFLLEGDOOHDPPTDSDSLSEDDQSGSYASTHSSDSEBEEBEEBBAAPGBOG 2760
Db 2701 DPGSLFLLEGDOOHDPPTDSDSLSEDDQSGSYASTHSSDSEBEEBEEBBAAPGBOG 2760
QY 2761 WDSLIGGARLPPLHSTPKDGGPGKAPWPGDFTTAKESGNGAPBEBRLRENGALSR 2820
Db 2761 WDSLIGGARLPPLHSTPKDGGPGKAPWPGDFTTAKESGNGAPBEBRLRENGALSR 2820

```

```

QY 2821 EGSLGFLPGSSAOPNR-----GILKKCLPTI 2847
Db 2821 EGSLGFLPGSSAOPNRHGMWGTSPCRAPLVSLIPLHMLPAPRPIGILKKCLPTI 2880
QY 2848 SEKSLRLPLECTSSSGSSASBSRGPPPPRPPROSLQOLNGVPIAMISIAQY 2907
Db 2881 SEKSLRLPLECTSSSGSSASBSRGPPPPRPPROSLQOLNGVPIAMISIAQY 2940
QY 2908 DEDSGSEFLFNFVFLH 2923
Db 2941 DEDSGSEFLFNFVFLH 2956

RESULT 8
US-09-737-149-25
; Sequence 25, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-737-149-25

Query Match 57.7%; Score 8974.5; DB 9; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

QY 1 MSPRATGVLPFPPPPPLLLLLLPPPLGQVQPCSLSGRGS-----S 48
Db 124 LRBSRGAELRSP-----AVRSVPLGDAL--CPRACGGAASLTSVLEAITNPPA 172
QY 49 GACAPM-----GWLCC--PSSASNLMTYSTRCDRAGTELTGHLVPHHDGLVWCPREBA 99
Db 173 CCCEPVAGCGRGRGICLRPGGSABLRVLCALGRAAGA-----VWV----- 213
QY 100 HPLPAPBPCWSCRLIGIIGHLSPOGKLTLPBHPCLKAPRLRCQSCKLAQAPGLRAG 159
Db 214 -----ELVYQATSGTPSSPSV-SPSL-----LNLQOP--RAG 243
QY 160 --ERSPEBSLGGRRKRNVTAPQOPPSYQATVPENOPAGTVASLRATIDPEGEAGRL 217
Db 244 VVRKS-----RKGTGSGTSPQPLPSTQVSFENRPGTAVTELRANDPBGDGRKS 296
QY 218 YTMALFDRSRNQPSLDPVTGAVTTABELDETKSTHVFRTAODHGMRRSALATLTI 277
Db 297 YGMELFDRSRNGYFLIDATGAVTTABELDETKSTHVKSAVDHGSRRSAAATYLTIV 356

```

Qy	278	VTPIIDNDPVPEOQYKESLRENLVEGCVLTVAATGDAPANNITLRLLEGSGSPS	337
Db	357	TVSDINDSPVPEOSEYERIKENLEVEGCVLTITATGDASNNMRRLLEGAG---	413
Qy	338	EVFEIDPSPGVIKRTGCPVDREEVESYOLTVESADCGPCGPPSTAAVFLSVEDNDNAP	397
Db	414	-VFEIDASGVVTRTAAYVDBREAAAYOLLVEANDCGRPGPLSAAVTHIIVEDENDYF	472
Qy	398	QPSERRVVOVREDYTPGAPVLKVTASDRDKGSNAVHYSIMSGNAQOPLYDACTGALD	457
Db	473	QPSERRVVOVPEDEVAVNTAVLRVQATDRDQONAAIHYSIVSGNLKCOFLYHLSGSLD	532
Qy	458	VVSPLDYETKREYTLKRAVADGGRPLSNVSGLYVOULINDNAPIVTSVPRQATYLES	517
Db	533	VINPUDFPAIKREYTLKRAODGRPLINSSGLSVSOVLVDNDNAPIFVSSPFOAAVLEN	592
Qy	518	VPLGLVTHVVAIDADAGDNARLEKLAGVGH-----DPEFTINNGTG	560
Db	593	VPLGHSVLHIGAVDADAGENARLOVLVDYASTIVGSSVDSENASAPDPPFOIHNSSG	652
Qy	561	WISVAELDREEDVYSFGVEARDHGTALITASASVTVLVDVNDNPTFOPEYVALN	620
Db	653	WITVCAELDREVEHYSFGVEAVDHGSPAMSSASVSTVLVDVNDNDPMFQVPEYELN	712
Qy	621	EDAAVNGSVTVVSAVDRDAHSYITQITSGNTRNPFSTISGSGGLVSLAPLDYKLERQ	680
Db	713	EDAAVSSVTLRADRDANSVITQITGNGTRNPFALSSOGGLITLALPLDYKOSRO	772
Qy	681	YLVAVTASDGTRODPAQIVNVTADNANTHRPVQSSHVTVNVNEDPAGTIVVLSATGED	740
Db	773	YLVAVTASDGTGRSHRAQVFINVTADNANTHRPVQSSHVTVVSEDEPVGISTATISATED	832
Qy	741	TGENARITYFMEDSIPQRIADDTGAVTTOAELDEDOVSYTLATARDNGIPQKSDITY	800
Db	833	TGENARITYVEDPVQPIRIDPDTGTITYTMELDYEDDOATTLATITADNGIPQKSDITS	892
Qy	801	LEILVYVNDNAPQPLRDSYQGSVYEDVPPTSVQISATDRDGLNGRVFTYFGSGDGG	860
Db	893	LEILILDNDNAPRLRDPYQGSVPEDAPPSTSVQVATDRDGLNGRLTYTFOGSGDGG	952
Qy	861	DGDFIVESTSGIVRLRLDRBNVAQVLRVAVDVKMP-PARTMETVTVVLDVNDMP	919
Db	953	DGDFIETSGVIRTORLDRBNVAVYLMALADVKGSPNPLSASVGIQVSVLDDNDMP	1012
Qy	920	VFBODEPDPVFEENSPIGLAAVARVATPDPDEGTNAQIMQIVKGNIPVFPOLDIPSGELT	979
Db	1013	VFBODELELTFEBENSFVGSVVARIRANDPDEGNQIITQIVKGNVPEVPOLDILSGDLR	1072
Qy	980	ALVLDVDEDPREYVLYVQATSAPLVSRATVHVLIDRNDNPPVLGNFEILFNNTYTNSS	1039
Db	1073	ALVLEIDPEVRDYMVLVQATSAPLVSRATVHILRLDQNDNPELPPDQILFNNTYTNKSN	1132
Qy	1040	SFPGGAGORVAHPDIDSLSLTVSEBENELSLVYLNASTGEKLSRALDNNRPLAAMS	1099
Db	1133	SFPGSVGIRIHAHPDIDSLSLTVLQNEJLSLLLDATBGLQISRLDNNRPLEALME	1192
Qy	1100	VLVSDGVHSAQCALRYTIIITDEKLTISITLRLDEMSPERFISPLGLFIOAVATLAT	1159
Db	1193	VSVSDGHSVTRALCTLRVITITDDMLTMSITVRLNMSQKFTSLSLPVBGAVALYST	1253
Qy	1160	PPDHVVVFNVRDTPAGGHILNLSVSGPPGPGGPPPLPSBDLQERLYIAMSLLTAI	1219
Db	1253	TKDIDIPVENIONDYV-SSNITVTFSLALPGTGRG--RFPSEBLOQOIVIANRILTTI	1309
Qy	1220	SAOSVLPDPDNDICARECENTMRCSVLRPPOSSAPFIASSSTLPRHHPVGGLACRCRPG	1279
Db	1310	SAORLPLPDDNDICARECENTMRCSVLRPPOSSAPFISSTVLPFPHIPITGLRCRCPG	1365
Qy	1280	FTGDIYCEFEVULCYSRPCGPHRCRSRREGVYTCIRDIDYTEHCVANSRGRTPGVCYN	1339
Db	1370	FTGDIYCEFEIDLCYSNPCGANGRCRSRGVYTCCEFBPFTGHECVANRSRCSAGVCYN	1429
Qy	1340	GGTCVNLIVGGKCDSPSGEKPQYQVYTTSPFPAHSFITERGLRQRFHTALSPATKE	1399

Dh	1430	GGTCVNLIIIGFHCVCPEGEVBNHYCEVSTRSPDQSFVTFERGRLORFHFVLSAFAFAD	1489
Qy	1400	RDGLLLYNGRNRNEGHADVAALEVIIEQVOLFPSABESLTVVSPFVPGVSDQOMHTVOLKY	1459
Dh	1490	RNALLLNGRPNRNEGHDIETAEIYERQOLFPSAGETITTTVPOVPGVSDGRMHSVLVOY	1549
Qy	1460	YNKPLIQLQOTGLPOGSPBOKYAAVVTVDGCTGVALARFSGVIGANYSCAAQGTGGSKSLDL	1519
Dh	1550	YKNKENIGHLGLPHGSPGEKVAVVTVDCCDAVAVHFGSVYGNYSCAAQGTGGSKSLDL	1609
Qy	1520	TGPIILIGVDPDPESPFPVRBRQFPGCBMRNTQVDSRHIDMADFIANNGTVPCCPAKAVCD	1579
Dh	1610	TGPIILIGVDPDPESPFPVRBRQFPGCBMRNTSIDRIYDMAAFIANNGTWGCASQRRFCD	1669
Qy	1580	SNTCHNGATCVNQMDASCECPLOFGSGKCAQOEANQOHFGSSLVAMHGSLPISQPMY	1639
Dh	1670	GTSQONGSTCVNKRNTYLCCEPLRFGKACBQAMPHQRTGSESVLMSDLDTITSIPWY	1729
Qy	1640	LSIMFTRQADGVLLQAITRGRSTITVQLREGHVMLSVEGTGLOASSLRLEPGRANDGW	1699
Dh	1730	LGLMFRTRKEDGVLMAYAGTSRSLHQLIINSYIRFEVSYGSPDVASMQLSKSRITDGM	1789
Qy	1700	HHAQALGASGSPH-----AIIISFDYGOQBARBENLPRHLGHLNSITVGGIPGPAGV	1754
Dh	1790	HHLLIEL-RSAKEBKDIKYLAVMTLVDGMOISTVQIOLNQLPGLMRTITIVIGVTEDEVSY	1848
Qy	1755	ARGRGGLQGVRAVDTEPEGVNSLDPSHGSLINVEQGSLLDPCDSNNCPANYSYNSNMDS	1814
Dh	1849	RHGRRGCQGVRRGEGISTNTAITAIIANNALAKRVXVDGCVDEPCASSPCPPHPRCDRWDS	1908
Qy	1815	YSCSDCPGYVDNCTVNCIDINPCEHOSVCTRKPSAPHGYTCCECPNVLGBCYCEFRIDOPC	1874
Dh	1909	YSCICDRGYFGKKCVADCLNPPCHVAAVCYRSPNTPRGVCEGSPGHYGQYCEKXKVDLPC	1968
Qy	1875	PRGWWGHPITGPCNCVDYSKGFDPDCNTSGBCCHKENHYRPPGSPYLLCDCTPYGSLR	1934
Dh	1969	PKGMWGPVCGPCGCAVSGFDPDCNKTNGCQCCKENYKRPADACLPCDCPFHSGHSR	2028
Qy	1935	VCDEBDQCCCKPEVIGRQCDRCNPNFAXTVINGCEVNNYSCPAETAGIWMPTRGLP	1994
Dh	2029	ACDMDTQCAKCEKGVIRQCNCRCONPFAEYTSIACEVYINGCPCPAFEGIWMPTRGOP	2088
Qy	1995	AAAPCPKGSFGYAVRHCDERHGMPLPVLFWCTSIYFSEBELGFAERLORNESGLDGSQOQ	2054
Dh	2089	AAVPCPFGSGVGNVHRHSGSEKGMPLPELFWCTSGSPFDLALNWKLNMRNTRBMDGNLSR	2148
Qy	2055	LALILRNATQHTAGYFGSDYKVAVOLATRLILAHSTORGFGLSATQDVHTENILARVGA	2114
Dh	2149	LAKALRNATQGNSTFLFENDVRYAYQLARILIOHRSRQGGFDLAIATREANFHEIDVHTGSA	2208
Qy	2115	LIDTPANRHEILIOETGCTAMLIQHEAYASALQOMRHTYLSPTITVPNIVISVRL	2174
Dh	2209	LLAPATASNEQIORSBAGAAQLRHEEAFVSNARVAKRTYLRPIVITANNITLADIF	2268
Qy	2175	DKGNFAGAKLPRYEALRGEQRPDLETTVIIPESVF-----RETPEVYV-----PAGP	2221
Dh	2269	DKNFTAQVPRPBDIOELPRELESVSPPATFTRPKCKGGEVYRLITMRKTPPLTAOP	2328
Qy	2222	GBAQEPBELARQRRHELSQGBAHSVITIYRTLAGLLPHNYDPPKRSLEVPKRPITNP	2281
Dh	2329	EPRAERETSSRRRRHDPDPCQAVALLVYRTLQGLLPHYDPRHRSRLPNRPVINTP	2388
Qy	2282	VVISVYHVDDELLPRALDKRVYVQOFLLEFEETKPICVPMNNSILVSGVGMASRQCEV	2341
Dh	2389	VSMANVSEGTPLSPSSIORPILVEFSLLETERSKVPCVPMNNSLDGCGVGMASAGCEL	2448
Qy	2342	VFRNESHVSCQCNHMTSFAVLMQVDSRRENEBILPLKTLVVALAGVTLAALLTFEFTLL	2401
Dh	2449	LSRRKRTVTCQCHSASCANTVMSISREHGEVLPKTIITAAISLISVALLVAFVILSV	2508
Qy	2402	RIIRSNOHGIRNVLTAALGAQVFLTLGINQADLPACTVIAIILHFLVLTCTSMALLEA	2461

```

Db      2509 RTLRNLSHSHKNI1AALFESOLI1FWGINOTENPFLCTVVAIILHYVSKGFAWTLVEN 2568
Qy      2462 LHLRYALTEYRDVNTGPMRFYMLGNGVPAFITGLAVGLDPRGYNDDPCWMLSYCTLM 2521
Db      2569 LHYRYMLTEFRNIDYMGRIYHVGWGIPIAVGLAGLDPQGYGNPFCWMLSDOTLW 2628
Qy      2522 SFAAPAFAVMSVFLYIIIAARASCAAOQGFKKGVSGIQPSFVALLSLASTWLLALL 2581
Db      2629 SFAAGVGTAVIIIVTVFVLSAKVSCQKHAYERKGVSWLRTAFILLVLTWTLGLL 2688
Qy      2582 SVNSDTLFFYLVATCNCTICGPPILFSLVYVLSKEVTKALK-LACSRKPSDPALTTKST- 2639
Db      2689 AVNSDTLSFYVLPAAASCLQGIIVLLFHCVAHREVRKHLAVLAGKQLQDLSATTRALT 2748
Qy      2640 LTSYNSCPSPYADG--RLYQPYGDSAGSLHSTRSGKSGPSYIPLLRBSALNPQGGP 2697
Db      2749 LTTSLNCNNTYSGPDMRLTALGEBSTASLDSTRDEVO-----KLVSQGPARG 2798
Qy      2698 GLADPGSLFL-EGODQOHDPTDSDSLSDQSGSYASTHSSDSEEEHEEEEAAP 2756
Db      2799 NHGEPDTSFIPRNSKKAHGPDSDDSELSL-DEHSSSYASTHSDSEDDGGAEDK----- 2853
Qy      2757 GEGGMSLGPAGARLPLHSTPK-DGPGGKAPWPD--FGTTAKG----- 2800
Db      2854 ---WNPAGPA-----HSTPADALANHVPAWPDSELAGSDSELDTEPHLYETKV 2903
Qy      2801 ---SSGNGAPEBERLRENGALSRGSLGPLPGSSAOP---HKGIKKKCL-PTIS 2848
Db      2904 SVELHQAQGNHCGDRSDSESVLAK-----PVAVLSQPOEORKILNKYTPPPPLP 2958
Qy      2849 EK--SSILRLPLEQCTGSSRGS--SASEGRGSGPP-----RPPROSLQOQLNGVMP 2898
Db      2959 EGPLKSLREKLADCEQSPSSRTSLGSGDVHATCVITIKTPRPERBHLNGV--- 3015
Qy      2899 AMSIKAGTVDEDSGSE 2915
Db      3016 AMNVRTGSAQANGSDSR 3032

RESULT 9
US-09-737-149-30
; Sequence 30, Application US/09737149
; Patent No. US2002007466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 30
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus

```

```

US-09-737-149-30
Query Match      57.7%; Score 8974.5; DB 9; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

Qy      1 MSRPATGVLPPEPPLLLLLLLLLLPPPLQVQGPCSISGRGSGS-----S 48
Db      124 LRSRARGELRSP-----AVRSVQLGDL--CPRAGGGAASLSVLEALTNPFA 172
Qy      49 GACAPM-----GMLC--PSSASNLMLYTSRCRDAGTELGVPHHDLRWCPSEBA 99
Db      173 CSCPVAGTGRGRRGPICLRPFGSAELRLVCLAGRAAG-----VNV----- 213
Qy      100 HILPLPAPBGCWSCRLGIGIGHLSPOGKTLIPREHPLCLAPRLACSCKLAQAGLRAG 159
Db      214 -----ELVIQNTSGTPSSPSV-SSSL--LNLSP--PAG 243
Qy      160 ---ERSPEESLGRARRKNVNTAPQOPPSYQATVPENQAGTVPASLRAIDPEBAGHLE 217
Db      244 VVRG-----RGTGSSSTSPQPLPSYQVSPENBAGTAVIELRAHDPBEGAGRLS 296
Qy      218 YTMALFDSRNSQFSLDPVTGAVTTABELREYKSTHFRVTAODHGMRRSALATLT 277
Db      297 YQMEALFERSNGYFLIDAATGAVTTASLDRETQTHVLSAVDHGSPRSAAVTLTV 356
Qy      278 LVTDTNHDVPVEQOEYKSLRENLGVGYVTVARATGDAPPNANILYRLLEGSGSPS 337
Db      357 TVSDTNHSPVEQSEYERIRENLGVGYVTVARATGDAPSNANMKRYLLEGSGS--- 413
Qy      338 EYFEIDPNSGIVRTSGPYVREBEVSQYLVEASDGRDPGPRSTTAAYFLSVEDNDNAP 397
Db      414 -VFBIIDASGVVTRAVVDREBAAYQLLVEANDGRNPGPLASATVHIVEDENDVY 472
Qy      398 QPSEKRYVQVEDYTPGAPVLRVTAORDKSNVAHYSINSNARQGYTIDAQTALD 457
Db      473 QPSEKRYVQVEDYVAVTVAVTVATDRDQONNAIHYSIVSGIKQGYLHYSLSGSLD 532
Qy      458 VVSPLDYETTKETLRVAADGGRPEPLNSGLVYVQVLDIDINDAPIFSTPFOATLES 517
Db      533 VTNPLDFEALIREYTLRIKADGGRPLNSSLVSVQVLDVNDNAPIFSSSFOALVEN 592
Qy      518 VPLGYLVHVOAIDADAGNARLEAYRLAGVH-----DPPFTINNGTG 560
Db      593 VPLGSHVLIQAVDADAGNARLQYRLVDTASTIYSGSSVDSBNPASAAPDFQJHNSG 652
Qy      561 WTSVAELDRREVYVSGVEARHDGTPALTPASVTVYDVNNNPFQPEVTVRN 620
Db      653 WITVCHELDRREVERHYSFGVEAVDHGSPMSSASVITVLDVNDNPFQPEVTVRLN 712
Qy      621 EDAAVGTSVVTVASVDRDAHVSITYQITSGNTRNFSITSGGGGLVSLALPLDYKLERQ 680
Db      713 EDAAVGSVTLIRADRDANSVITYQLTGANTRNFPALSSGGGLITALLDLKQERQ 772
Qy      681 YVLAATASDGTRODTAQIVNVNTDANTRPVFOSSHVTVNNEBRPAGTTVVLISATBED 740
Db      773 YVLAATASDGTSHTAQVFINVTDANTRPVQSHYTVSVEBRPVGHSIATISATBED 832
Qy      741 TGENNARITFMEDSLPQRIADTGAVTQALDVEDQVSYLATAIDNNGIPQSDPTLY 800
Db      833 TGENNARITVLEDPVQFQRIIDPTGTITWELDEDAAYLATIAQDNGIPQSDPTLS 892
Qy      801 LEILVNDVNDNAPQFLRDSYQGVYVEDVPFYSVLQISATDRSGLNGVFTYFGGDDG 860
Db      893 LEILILANDNAPRFLRDFQYGSVEDAPPSVSVQVSTATDSDGPNGLLTFPGGDDG 952
Qy      861 DGDFTVESTSGIVRTLRKLDRENVAAQYLRAYAVDKMP-PARTMEVTVTVLVNDNPP 919
Db      953 DGDFTYIEPTSGVIRQRLDRENVAVVYMLAAVDRGSPNPLASVGIQVSVLNDNPP 1012
Qy      920 VFEODEPVDVFEVENSPIGLAVARVATPDDEGTNAQIMVQIYEGNIPYFOLDIFSGLT 979
Db      1013 VFEODELIFVEKNSPVGSVVARIRANDPDEGPNMOIITQIVEGNVFVYFOLDLSGLR 1072

```

QY 980 ALVLDYEDPREVYVITQATSAPLVSRATVHVLDRNDNPVLGNFELLFNNYVTRSS 1039  
 DB 1073 ALVLEDFEVRBDMLVQATSAPLVSRATVHVLDRNDNPVELPQFQLFNNYVTRKSN 1132  
 QY 1040 SPFGGALGRVPAHDPIQSLTYSFPERGHELSVLTMASTGELKSRALDNNRPLEAMS 1099  
 DB 1133 SPFGGALGRVPAHDPIQSLTYSFPERGHELSVLTMASTGELKSRALDNNRPLEAMS 1192  
 QY 1100 VLVSDBVSHVTAQALRVITITDEMTHSITLLEJMSPERFLSPLGLFQVAATLAT 1159  
 DB 1193 VASVDSGHSVTLALCTARVITITDDMLTNSITVRLNMSQKFLSPLSFGVAVLST 1252  
 QY 1160 PPDHVVVENVQDTPAPGHIILNVSLSVQPPGPGGPPFLPSBDILOERLYNRSILTAI 1219  
 DB 1253 TTDQDIFVFNIOQDTPV-SSNIILNVTSALLPGSTRG--RFPSPEDLOEQIYNRLTLTTI 1309  
 QY 1220 SAORVLPFDNITCLAREPCENYRCSVTLRFDSSAPPLASSVLPRIHVGRLRCRCPG 1279  
 DB 1310 SAORVLPFDNITCLAREPCENYRCSVTLRFDSSAPPLASSVLPRIHVGRLRCRCPG 1369  
 QY 1280 FTGDYCEYEDVLCYRNPCCPHGRCSRREGGYTCLCRDGYTGEHCEVSARSGRCTPGVCN 1339  
 DB 1370 FTGDYCEYEDVLCYRNPCCPHGRCSRREGGYTCECEDEFTGHCQVNVTSGRCSAGVCN 1429  
 QY 1340 GGTCTNVLVGGFKCDSPGDFEKEYCQVTTSPHASFITTFGLRORFHTLSPATKE 1399  
 DB 1430 GGTCTNVLVGGFKCDSPGDFEKEYCQVTTSPHASFITTFGLRORFHTLSPATKE 1489  
 QY 1400 RDGLLLNGRFRNEKDFVALLEVOVOLTFPSAGSTTTVSPFVGGVSDQGMHVLQKY 1459  
 DB 1490 RRLALLNGRFRNEKDFVALLEVOVOLTFPSAGSTTTVSPFVGGVSDQGMHVLQKY 1549  
 QY 1460 YNKPLLGQTLPOGSEBOKVAVVTVDCDCTVALRFGSVLTAGVSCAAGTQSGSKSLDL 1519  
 DB 1550 YNKPLLGQTLPOGSEBOKVAVVTVDCDCAAVAHVFGSVYVSCAAGTQSGSKSLDL 1609  
 QY 1520 TGPILLGGVPLDPESEPVARMROFVGCNRLQVDSRHIDMADFIANNGVPGCPAKKAVCD 1579  
 DB 1610 TGPILLGGVPLDPESEPVARMROFVGCNRLQVDSRHIDMADFIANNGVPGCPAKKAVCD 1669  
 QY 1580 SNTCHNGGTCTVNMOMAFSCCEPLGFGGKSCAQBAMPOHFLGSSIVAMHGLSLPISQMY 1639  
 DB 1670 GTSCONGGTCTVNMOMAFSCCEPLGFGGKSCAQBAMPOHFLGSSIVAMHGLSLPISQMY 1729  
 QY 1640 LSLMERTQADGVLLQATITRGRSTITLQIREGVMLSVYEGTGLQASLRLERGRANDGM 1699  
 DB 1730 LGLMERTRBDGVLMBAITAGTSSRLHLQILNSYIRFVSYGSPDVASMLSTSRITTDGM 1789  
 QY 1700 HHAQLALGASGGRH-----AIIISPDYGOQRAEGLGRLHGLHLSNITVGGIPEGAGV 1754  
 DB 1790 HHLLEL-LSAKEGKDIKYLAVMTLDYGMDSITVQIGNQLPGLKRRITVIGVTEDKVSV 1848  
 QY 1755 ARGFGGICGVAVSPDREGVNSLDPSSHGESSINVEGCSLPBPCDNNPCANSYCSNMDS 1814  
 DB 1849 RHGFPGCQGVMMGFTSTNIALTNMMDALKVAKGQGVDEDCASSPCPHNRCADTIDS 1908  
 QY 1815 YSCSCDPGYGDNCTNVCDLNPCEHOSVCTRKPSAPHGYTCECPNVLGAPYCETRIDPC 1874  
 DB 1909 YSCICDRGYFGKKCTVDACILNCPKVAACVRSPTNPRGSCGCGHGYQKENVVDLPC 1966  
 QY 1875 PRGMWGHPTGFCPCNCDVSKGPPDCKNTSGECHKENHTRPPGSPCTLCDGYPTGSLSR 1934  
 DB 1969 PRGMWGNPVCPCCHCAVSGPDPCKNTKMGQCKENYKPPADQACLPCCFPHSHSR 2028  
 QY 1935 VCDPBDGOCPCRGVILGRQCDRCNDPFAVTTNGCEVNVYDSCRRALIBAGIMWRTRFGLP 1994  
 DB 2029 ACDMDTGOCACRGVILGRQCDRCNDPFAVTTNGCEVNVYDSCRRALIBAGIMWRTRFGLP 2088  
 QY 1995 AAAPCPKGSFGTAVHCDERHGMPLPNTFNCTISFTEBLKGFARLQHNESGLDSGRSQ 2054  
 DB 2089 AAAPCPKGSFGTAVHCDERHGMPLPNTFNCTISFTEBLKGFARLQHNESGLDSGRSQ 2148

QY 2055 IALLIRNATQHTAGYFGSDVAVQVQATRLLAHSTORCGFGLSATQDVHFTENILRVGSA 2114  
 DB 2149 IAKALRNATQGSTLFGNDVATAVQLLARILOHESROGFDLAAREANFHEVDHGTSA 2208  
 QY 2115 ILDTANKHMLIOOTEGGTAWLLOHYEAVASALAAQNRHLYLSPTVTPNIVSVRL 2174  
 DB 2209 LIAPIEASWEQIOESKGAQQLRHFEAFPSNVARNKRRTYLRPFVITAMILLAVIIF 2268  
 QY 2175 DKGNFAGAKLPRYEALRGEPDLETTVILPESV-----RETPVVR-----PAGE 2221  
 DB 2269 DKANTGQVRFREDIOBELRELESVSFPADITKPEPKKGEPVRLTNKRTTLPQ 2328  
 QY 2222 GEAQPEELARRQRHPELQGEAVASVITYRTLAGLLPHNYDPDKSLAVPKRPIINTP 2281  
 DB 2329 EPRARETSSRRRRHPEPCQFAVALVIVITIGQLLPEHYDPDRSLRPNRVINTP 2388  
 QY 2282 VVSIIVHDEBELRALDKPTVQPRILETEERKPCIVFNNHSLVSGTSGMSARGCEV 2341  
 DB 2389 VVSANVYSEGTPLPSLQRPILVERSLLETERSKPVCVFNHSLDTGTGMSAKGCEL 2448  
 QY 2342 VFRNESHVSCCNHMTSPAVLMDVSRRENGEILPLKTLTYVALGVTALALITPFLTL 2401  
 DB 2449 LSRRTHTVTCQCSHSASACAVLMDISRRHGVLPKXITTYALSLVALVAVFLSLV 2508  
 QY 2402 RILSNQHGIRRNITPAIGLAQLVFLGINDADLPACTVAILLHFLYLCFPMALLIA 2461  
 DB 2509 RILSNHLSIHKNTIALLFPFQLIFMVGINTENPFLCTVVAIILHYSWGTFAVTLVEN 2568  
 QY 2462 LHLVRLTEVDVNTGPRRFFYTMGKVPAPITGLAVGLDDEGVNPPCMLSTYDTLIW 2521  
 DB 2569 LHVTRMLTEVNNDIDGPRRFFYHVGWGPALVTLGAVGLDQGVNPPCMLSTYDTLIW 2628  
 QY 2522 SPAGVFAVAVMSVFLYIILARASCAAROGFEKKGPVSGLOPFAVLLLSATMTLAL 2581  
 DB 2629 SPAGVGVITIIINVIYVLSKVCQKRNHYERKGVSMKRTAPLILLVATLWLGTL 2688  
 QY 2582 SVNSDTLLFHYLPATNCIOGPFILSVVLSKEVRALK-LACSRKSPDPALTYKST- 2639  
 DB 2689 AVNSDTLSFHYLPFAFCLQGIPTVLPFCVHRRVRKRLRAVLAKKLQLDSDATVTRTL 2748  
 QY 2640 LTSYVNCPSYADG--RLVQPYGDSAGSIHSTSGSGSQSPFYITPLAREBALNPGQRP 2697  
 DB 2749 LTRSLNCNNTYSBPDMLRTALGSTASLSDSTYDBGVQ-----KLSVSSGPARG 2798  
 QY 2698 GLADPGSLFL-EGDOQDHPDTPDSDSLSEDDGSGSVASTHSDSEEBEESKEAFA 2756  
 DB 2799 NHGPRDTSFTRNSKKAHGPDSDSLSL-DEHSSSYASHTSDSDGGEADK----- 2853  
 QY 2757 GEQGDLSLPGFARLPLHSTPK-DGGRPGKAPWPCD--FGTTAKG----- 2800  
 DB 2854 -----WNPAGCPA-----HSTPKADALANHVAPGMPDESILAGSDSEEDTEPHLKVETKV 2903  
 QY 2801 -----SGNGAPERLEKRGDALSREGSLGRLPGSSAOP--HKGILKKCC--PTIS 2848  
 DB 2904 SVBLHROQGHGCDRPSDPESGVIAK-----PAVLSSQPOBORKGILNKKVTPPLP 2958  
 QY 2849 EK--SSLRLPLBOCTSSRGS--SASESGRGAPP-----RPPROSLQOJLGVMP 2898  
 DB 2959 EQPLKSLRREKLADCBGSPSSRTSSLSGSDGVATQCVITIKTPRHPGHEHANGV--- 3015  
 QY 2899 AMSIKAGTVDEDSGSR 2915  
 DB 3016 AMNVRTGSAQANGSDSE 3032  
 RESULT 10  
 US-10-131-409-70  
 ; Sequence 70, Application US/10131409  
 ; Publication No. US20030199465A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Malayankar et al.  
 ; TITLE OF INVENTION: No. US20030199465A1el polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 15966-675CIP1CON1



```

CURRENT APPLICATION NUMBER: US/10/131,409
CURRENT FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: 09/898,954
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: 60/182,733
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 60/182,724
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 60/183,896
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: 60/184,497
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/224,157
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/184,482
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/184,744
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/197,083
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/233,405
PRIOR FILING DATE: 2000-09-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 135
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 3034
TYPE: PR
ORGANISM: Mus musculus
US-10-131-409-70

```

```

Query Match      57.7%; Score 8974.5; DB 12; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

```

```

1 MRSPTAGVPLPTPPPLLLLLLLPPLLDGVDGFCRSIGRSGS-----S 48
124 LRSASGAEIRSP-----AVRSVFGIDAL--CFPAAGGGAALSTSVLAITNTPA 172
49 GACAPM-----GWL- -PSSASNLMLYTRCRDAGTELTHLVPHHDLRWCPESBA 99
173 GSCPPVAGTCRCRGPICLRPGSAELRLVCAIGRAAGA-----VWV----- 213
100 HILPAPACGSCWSCRLLGIGHLSPGKLTLPBEPCLAKAPRLRCQSCLAQAPGLRAG 159
214 -----ELVIGATSGTSPSSPSV--SPSL-----LNLSPR--RAG 243
160 --ERSPEBSLGGRRKRWNTAPQOPPSYQATYPENOPACTPVASLRALIDPDGEAGRIE 217
244 VVRS-----RRGTGSSSTSPQPLPSYQVSVENEPACTAVIELRAHDPDGDGRLS 296
218 YTMADLPDSRNPFSLDPVYGAVTTAELDRKTSKTHVFRVTAODHGMRRSALATLT 277
297 YQEBALPDSRNGYFLIDAATGAVTTARSLDRETKOTHLKSAVDHGSRRSAATYTLV 356
278 LVTDTMDHDPFROQETKSLRENLVEGYEVLTVRATDGPAPNANILYRLLSGSGSBS 337
357 TVSDTNDHSPVFPQSEYRERIRENLVEGYEVLTVRATDGPAPNANIRYRLLSGAGG--- 413
338 EVREIDPRSGVITRGVNDREBSYQLTYSADQGRDPBPRSTTAANFLSVEDDNDNA 397
414 -VEIDRSGSVTRAVRAVDREBAEYQLLVEANDQGNPGLSASAVHIVDEBNDNP 472
398 QFSEKRVVQVREDVPGAPVLRTVYASDRDGSNAVVHYISMGNGARQGYFLDAQGTALD 457
473 QFSEKRVVQVREDVAVNTAVLAVQADRDQGNAAHYISVSGNLKQGYFLLSGSLD 532
458 VVSLDYEETKTYLRAVDGGRPPISNVGLVTVVDINDNAPITVSTPQATVLES 517
533 VINPLDPEALREYLRIRKADGGPPLINSGLVSYQVLDVNDNAPITFVSSPQAAVLBN 592
518 VPLGYLVLYOALIDADGNARLELYRAGVH-----DFPTINGTG 560
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

593 VELGSHVLIHQAVDADAGENARLQYRLVDTASTIVGSSVDSNPASAPDFPQIHSSG 652
561 WISVAELDRREVDYSGVEARHDGTALTSASVSUTVLDVNNNPFTQPEYVPLN 620
653 WTTVCAELDRREVEHYSFGEAVDHSPPMSSASVITVLDVNDNDPFTQPEYELRLN 712
621 EDAAVGSVTVASADRDASHVITYQITSGNRNRFSTISOGGGLVSLALPLDKLEBO 680
713 EDAAVGSVTLTRADRDANSVITYQITGNTNRNRFALSSOGGGLTALPLDKLEBO 772
681 YTLAVTASDGTRODTAQIVNVNTDANTRPVFOSSHVTYVNNEDRPACTVVLISATBED 740
773 YTLAVTASDGTSHTAQVFINVTDANTRPVFOSSHVTYVSEDRPVGTSIATISATBED 832
741 TGNARITYEMEDSLIPORIDADGTAVTQAEVDYEDVSYTLATDANGIPQSDPT 800
833 TGNARITYVEDPVQPRIDPDGTITYMELDVEDQALYATLAQDNGIPQSDPT 892
801 LEILVNDVNDNAPQELRDSYQGSVEYEDVPFTSVLQISATDRDGLANGVFTYPCGDDG 860
893 LEILVNDVNDNAPRFLRDPYQGSVEYEDVPFTSVLQISATDRDGLANGVFTYPCGDDG 952
861 DGDPIVESTGTVTRRLRDRENVAQYLRAYAVDKMP-PARTMEVTVYLVNDNMP 919
953 DGDPIVESTGTVTRRLRDRENVAVNMLALAVDRGSPNPLASAVGIQVSVLTDINDNP 1012
920 VEGODEPQVFEENSPIGLAVAVTATPDGEGTNAQVQIYEGNIPFVOLDISGELT 979
1013 VEKDELEFVENSVPQSVVARIAPNDPDSFNQIITYQIYEGVPEVFDLDSGDDR 1072
980 ALVDYEDREPEYVAVIQAATSAFLVSRATVHRLDRDNDPVLGNFLLPNNTYNS 1039
1073 ALVEIDFEPDRDYMVLVQATSAFLVSRATVHRLDQDNDPVLGNFLLPNNTYNS 1132
1040 SFPGALIGRVAHDIDISLTYSPERGENSLVILMASTGELKLSRALDNRPEALMS 1099
1133 SFPSGVIGIRPAHDIDISLTYSPERGENSLVILMASTGELKLSRALDNRPEALMS 1192
1100 VLVSDGVSVTAQCALRTYITIDEMLTHSITRLTLDMSPERLSPILGLFIQAVATLAT 1159
1193 VSVSDGHSVTRALCTLRATYITIDMLTNSITVRLNMQEKLSLSTFVSGVATVYST 1252
1160 PPDHVVVNVQRTDAPGHHILNVSLSVQPPGSGPPPLSEDLQERLYNLSLTXI 1219
1253 TKDDIFVFNQNDTVV-SSNLTAVTFSLALPGTARG--RFPSESDLOBOIYANRLTYI 1309
1220 SAORVLPEDNITCLABPCENYKCVSLRPSASAFIASSSTLPRIPHVGLRORCPG 1279
1310 SAORVLPEDNITCLABPCENYKCVSLRPSASAFIASSSTLPRIPHVGLRORCPG 1369
1280 FTGDYCETEVDCYGRPCGPHRCRSRGGYTCRCRDGYTGHCHEVSARSAGRCTGVCKN 1339
1370 FTGDYCETEVDCYGRPCGPHRCRSRGGYTCRCRDGYTGHCHEVSARSAGRCTGVCKN 1429
1340 GGTGVNLLVGFGKCDPESGDFEKPQYQVTTNSFPASHPTTFGRRLORHFTLSPATKE 1399
1430 GGTGVNLLVGFGKCDPESGDFEKPQYQVTTNSFPASHPTTFGRRLORHFTLSPATKE 1489
1400 RGLLLYNGRFBKBDPVALYIOBOYOLTFBAGSSTTVSPFVGVGSDGDMHTVQLKY 1459
1490 RGLLLYNGRFBKBDPVALYIOBOYOLTFBAGSSTTVSPFVGVGSDGDMHTVQLKY 1549
1460 YNKPPLGOTGLPQGSBOKVAVVYTDGCDTVGALRFGSVLANSYCAAOGTGGSKSLDL 1519
1550 YNKPPLGOTGLPQGSBOKVAVVYTDGCDTVGALRFGSVLANSYCAAOGTGGSKSLDL 1609
1520 TGPILLGIVPDLPEFPARMQPVGCNRLQVDSHIMADITANGVTPGPAKKNVD 1579
1610 TGPILLGIVPDLPEFPARMQPVGCNRLQVDSHIMADITANGVTPGPAKKNVD 1669
1580 SNTCHNGGTQVNMQAFSCCEPLGFGKSCAEMANPQHFLLSSIVAMHGLSLPTISQPY 1639
1670 GTSQNGGTQVNMQAFSCCEPLGFGKSCAEMANPQHFLLSSIVAMHGLSLPTISQPY 1729

```

```

QY 1640 LSLMPTROADVLLQAITRGSTITLQREGHVMSVEGTGLQASLSLREGRANDGDW 1699
DB 1730 LGLMPTREDECVLMEATAGTSSRLHLQILNSYIREVEYVSPSDVSMQLSKSRITDGM 1789
QY 1700 HHAOLALGASGGRH-----ALISPDYGOORABGNLGPRLHGLHLSNITVGGIPBAGV 1754
DB 1790 HHLLEL-RSABEKGDIKYLAVWTLVDYGMDOSTVOQNLPGIKMRTIYIGVETEDKVS 1848
QY 1755 ARGFRCLOGVVSDTPREGVNSLDPHSGSINVEQCSLPDPCDSNPCCANGYCSNDMS 1814
DB 1849 RRGFGMCGVMGERTSTNIALTNMMDALKAVKOCVDVEDCASPCCPHPCADTMD 1908
QY 1815 YSCSCDPGYGDNCTNVCDLNPCEHOSVCTRRKPSAHHGYTCBPVNYLGPYCETRIDPC 1874
DB 1909 YSCIDRGYFGKKCVADACLNPCKHVAACVRSBNTPRGYSCEGPHGYOCENKVDJPC 1968
QY 1875 PRGMWGHPTGCPGNDVSKGPDCKNTSGECHKENHRRPGRSPCLLDCCYPGSLSR 1934
DB 1969 PRGMWGNPVCGPCHCAVSGQSPDCKNTNGCQCCKENYKPPADACLPCCDFPHGSHR 2028
QY 1935 VCDPEIDGCPCKRGVIGRCDCRCDNPFATVNTGCVNDSCPRAIEAGIMPTRTGFLP 1994
DB 2029 ACMDOTGCACCKRGVIGRCQRCNCDNPFATVNTSLGCEVYINGCPRAEAGIMPTRTGQP 2088
QY 1995 AAAPCPKSGFTAVHDCDEHRCMLPNNLFCNTSITPFSBLKGAERLQENESGLDGRSQ 2054
DB 2089 AAAPCPKSGVNAVHDCSEKGMPLPELFCNTSGSFVDLKALNEKLNREHETMDGNSLR 2148
QY 2055 LALLRNTOHTAGYFGSDVYKAYOGLAFTLLAHSTQRFGLSATQDVHTFNLLRVGSA 2114
DB 2149 LAKALRNNTQNSTLFGNDVRYAYOLLAIILOHESROQFDLAAREANFHDVHTGSA 2208
QY 2115 LLDITANKHMBELIOOTEGTAMLOHYEAYASALAONMHTYLSPTIYTPNIVISVRL 2174
DB 2209 LLAPEATASWEOIQSEBGAOLHHEFYFSNVANRVTTLRPFVITVAMMILAVIDIF 2268
QY 2175 DKGNFAGAKLPRYEALRGEOPLDLETTVLPSVF-----RTPPYVR-----PAGP 2221
DB 2269 DKLNTGQVPRFEDIOELPRELESVSFPADTFKPEKKGPRVRLNRRRTBLAQP 2328
QY 2222 GEAPREBELARORHPELSGEAVASYIYRTLAGLLPHNDPDKRSIRVPRKLTIMP 2281
DB 2329 EPPARETSSSRHRHPEPGFAVALVITYLQOLPEHDPDRSLRTPNRRVINTP 2388
QY 2282 VVSIHNDDEBLIPALDKPVTVQRLLETERTKPICVFNMHSILVSGTSGMSARGEV 2341
DB 2389 VVSANVYSEGTLPESLQRPILVERSLIETBERSKPCVCFNMHSIDTGTSGMSAKGCEL 2448
QY 2342 VFRNESHVSCQCNMTSPAVLMDVSRRENGETLPIKTLTYVALGVTLAALLTFPFTLL 2401
DB 2449 LSRNTHVTCQCSHSCAVLMDISRREHGEVLPIKITVYALSLVALVAFTLLSLV 2508
QY 2402 RLIRBNHGIRRNLAALGLAOLVFLGIGNOADLPACVYVAILHPIYLCFSPALLIEA 2461
DB 2509 RLIRBNHLSIHKNLLAALPFSOLIFWVGINQTEHPCLTVVAILHAYVSMGFPAATLVEN 2568
QY 2462 LHLVYALTEVDVNTGPMRFYMLGMVPAFTTGLAVGLDPEGYGNPFCMLSTVDTLIM 2521
DB 2569 LHVYMLTEVNNDIDGPMRFYHVGMPALVYTGAVGLDPEGYGNPFCMLSTVDTLIM 2628
QY 2522 SPAGVAVAVMSVFLYILABASCAAOQGFEEKKPVSGLOPSAVILLLSATWLLAL 2581
DB 2629 SPAGVGVVLIINTVIFVLASAVSCORKHYYERKGVVSMRTAFLLILLVATWLLGLL 2688
QY 2582 SVNSDPLLPHYLFAONCICQGFIFLASYVLSKVRKALK-LACRKSXSPBALTTKST- 2639
DB 2689 ANSDTLLSFHYLPAFSCLOGLFVLLFHCVAHREVRKHLRAVLAGKQLDSDSATRTVL 2748
QY 2640 LTSYNCSPPYADG--RLYQPYGDSAGSLHSTSRGKSQPSYIPLLEBSALNPGQPP 2697
DB 2749 LTRSLNCONNTYSEBGMRLTALGESSTASLDSTTRBGVQ-----KLSVSGSPARG 2798

```

```

QY 2698 GLDGFSLFL-EGDOQDHPDPTSDSDLSLBDQSGSYASTHSDSEEBEEREAAP 2756
DB 2799 NHGEPDTSFTRPNSKKAIGPDSDSLSL-DEHSSYASSHTSDSDDGGAEDK----- 2853
QY 2757 GEQGWDSLGRGAERLPLHSTPK-DGGRPGKAPPGD--FGTKKE----- 2800
DB 2854 ----WNPAGPA-----HSTPKADALANHVAGPDSLSLSDSEELDTEPHLKEVKV 2903
QY 2801 -----SSGNAPERERENGDALSRGSLGPLGSSAOP-----HGILKKKCL--PTIS 2848
DB 2904 SVELHROAQNHGCDRPSDPPSGVLAK-----PAVAVSSQFOEQRKGLIKKVTYPPPLP 2958
QY 2849 EK--SSILRLPLEOCTGSSRGS-SASGSRGAPP-----RPPRQSLQOLNGVPI 2898
DB 2959 EQLPKSLREKLADCEGPTSSRTSSLSGSGVHATDCVITIKTPRREGRHNGV---- 3015
QY 2899 AMSIKAGVDEDSGSE 2915
DB 3016 AMNVRTGSAQANGSDSE 3032

```

```

RESULT 11
US-10-120-801-52
; Sequence 52, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkels, Richard
; APPLICANT: Padigaru, Murailidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Foad
; APPLICANT: Topper, James N.
; APPLICANT: Malysankar, Uriel
; APPLICANT: Maeserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/286334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: mouse
US-10-120-801-52

Query Match 57.7%; Score 8974.5; DB 12; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

QY 1 MRSPTGVPLPTPPPLLLLLLLLLPPPLGDQVCPRSLSGRGSG-----S 48
DB 124 LKRSARGAELRSP-----AVRSVFGJGDL--CFPAAGGASLSTVLAITNPPA 172

```

QY 49 GACAPM-----GMLC--PSSASNLMLYTSRCRDAGTELTHLVPBHDLRWCPESBA 99  
 Db 173 CSCPAPAGVGCRCRGPICLAPGSSAEHLVICALGAAAGA-----VWV----- 213  
 QY 100 H1PLPAPBGCPCMSRLIGIHLSPQKLTLPBEPCLKAPRLRCSCKLAQAPGLRAG 159  
 Db 214 -----ELV1QATSGTSPSV-SPSL-----INTLSQP---RAG 243  
 QY 160 --ERSPEBSIGRRKNVNTAPOPPOPSYQATYPENOPACTPVASLRALDPDGEAGLE 217  
 Db 244 VVARS-----RRGTSSSTSPQPLPSYQVSVBENEPAGTAV1ELRADPDEGDGRIS 296  
 QY 218 YTMADLPDSRNOFSLDPVTGAVTAAEBLDRKTHYVRVYAADHGMRRSALATLT 277  
 Db 297 YQMBALPDRSKNYFLIDATGAVTTARSLDRKTHYLRKYSADHGSRRSAAVTLV 356  
 QY 278 LVTYNDHDPVFPQOEKESLRNLEVEVLTVRAITDGDAPPANILYRLLEGSGSBS 337  
 Db 357 TVSDTNDHSPVFPQSEYRERIRENLEVEVLTIRATDGDAPPANMRYRLLEGAGG--- 413  
 QY 338 EYVEIDPDRSVITRGPVDRBEVESYQLTVEASDQGRDPBRTTAAPLVSDEDDNAP 397  
 Db 414 -VEIDARSGVATRAVVDREBAEYQLVLEANDQGRNPGPLSASATVHIVDEDDNAP 472  
 QY 398 OFSEKAYVVOVREDVTPGAPVLRVTASDRDKGSNAVHVSIMSGNRGOPYLDAOTGALD 457  
 Db 473 QPSEKAYVVOVREDVAVNTRAVLAVQATDRQGNMAIHISIVSGNLKQPYLHLSGSLD 532  
 QY 458 VVSPIDYETTKETYLVRADQGRPLSNVSLVTVQVLDINDNAP1FVSTPQATVLES 517  
 Db 533 VINPLDFEALREYTLRIKQDQGRPLINSGLVSVQVLDVNDNAP1FVSSPQAVLEN 592  
 QY 518 VPLGYVTVHQAIDADAGDNARLEVRLAGVH-----DRETTNNGG 560  
 Db 593 VPLGSHVTLHQAADADAGENARLQYRLVDTASTTVGGSSVDSNPNASAPRPPQIHNSG 652  
 QY 561 WISVALELDEEVEFYSGVEARDHGTPALTASASVTVLDVNDNPP7TQCEYTVRLN 620  
 Db 653 WITVCELEDEEVEHYSFGVEADHGSBAMSSASISITVLDVNDNPPMTQVYELRLN 712  
 QY 621 EDAVGTSVTVSAVDRDAHSVITQYTGNTNRPSTISQSGGLVSLALPDYCLERO 660  
 Db 713 EDAVSSSVTLTARBRDANSVITQYTGNTNRPSTISQSGGLVSLALPDYCLERO 772  
 QY 681 YVLAIVTASDGTRODTQIVVNTDANTHRPVFQSSHVTAVVNBDRPAGTVVLISATDE 740  
 Db 773 YVLAIVTASDGTSHTAQVFINVTDANTHRPVFQSSHVTASVSEDRPVGISITISATDE 832  
 QY 741 TGENARITVEMEDSIPOFRIDADTGAVTTOAELEDEDOVSYTLATARDNGIFOKSDTY 800  
 Db 833 TGENARITVLEBVPQFRIDPDTGITYMTLEDEDOVSYTLATARDNGIFOKSDTY 892  
 QY 801 LELIIVDNDNAPQFLDYSQGYEDVEPPTSYLQISATDRSGLANGVFTTFOGGDQ 860  
 Db 893 LELIIVDNDNAPFLDYSQGYEDVEPPTSYLQISATDRSGLANGVFTTFOGGDQ 952  
 QY 861 DADPVESTGIVTTLRLRENVAVVLAAYADKMP--PARTPMEVTVTVLDVNDNP 919  
 Db 953 DADPVESTGIVTTLRLRENVAVVLAAYADKMP--PARTPMEVTVTVLDVNDNP 1012  
 QY 920 VFEODEFVVEENSPIGLAVARVATADPDEGTNAQIMYOIVBGN1PEVFOLD1FSGELT 979  
 Db 1013 VFEODEFLELFEVEENSPVGVARIRANDPDEGTNAQIITQIVEGNVFVFOLD1FSGELT 1072  
 QY 980 ALVUDYDEDEPEVTVLQATSPVSRATVHVLDPDNDPVLGNBELL1PNNYTRRS 1039  
 Db 1073 ALVBLDEVRDNYLVVQATSPVSRATVHVLDPDNDPPEL1PNNYTRRS 1132  
 QY 1040 SFPQGAIGRVAPHDPSIDSLTYSFPERGNEISLVLNASTGELSLBALNNRPLEAIMS 1099  
 Db 1133 SFPQGVIGRIPAHDPSIDSLNTYFLOGNEISLVLNASTGELSLBALNNRPLEAIMS 1192

QY 1100 VLVSDEVHVTQAOCALRVITITDEMLTSHITLRLDEDMSPERFLPDLGLFOAVATLAT 1159  
 Db 1193 VVSUHGHSVTRALCTLRATITITDDMLTSHITLRLDEDMSPERFLPDLGLFOAVATLAT 1252  
 QY 1160 PPDHVVENVORJDTAPGHIILNLSVGOQPPGCGPPLSEDLQERLYNRSALTAI 1219  
 Db 1253 TKDITFVFNITQNDTV--SSNITLNTVPSALLPGTGRG--RFPSESDLOEQIYNRTLTIT 1309  
 QY 1220 SAORVLPDDNCLAREBENTWCVSVLRFPSSAPF1SSVTLPRH1PVVGLRRCRCPG 1279  
 Db 1310 SAORVLPDDNCLAREBENTWCVSVLRFPSSAPF1SSVTLPRH1PVVGLRRCRCPG 1369  
 QY 1280 FTGDYCEVEVDLCYSRPGPHRCRSREGYTCLCRDGYTGHEHCVSARSAGCTGVCKN 1339  
 Db 1370 FTGDYCEVEVDLCYSRPGPHRCRSREGYTCLCRDGYTGHEHCVSARSAGCTGVCKN 1429  
 QY 1340 GGTGVNLLVGGFKCDPSPGDEPKPYQVYTTNSFPFASHTTFRGLRQRFHTALSFAIYE 1399  
 Db 1430 GGTGVNLLVGGFKCDPSPGDEPKPYQVYTTNSFPFASHTTFRGLRQRFHTALSFAIYE 1489  
 QY 1400 RDGLLYNGRFRBKHDPALEIOLQVOLTSPSAGSTTVSPFVRGVSDGOMHTVQLKY 1459  
 Db 1490 RNALLVNGRFRBKHDPALEIOLQVOLTSPSAGSTTVSPFVRGVSDGOMHTVQLKY 1549  
 QY 1460 YNKPILLQGTGLPQGBSEOKVAVTVVDCDVTGVALRFGSVLGNYSQAOGTGSGSKSLDL 1519  
 Db 1550 YNKPINIGHLGLPHGSGERKAVVTVDDCAVAVHFGSVGNYSQAOGTGSGSKSLDL 1609  
 QY 1520 TGPILLGGVPLDPSFPVPMROFVGCNRLOVDSHIMADFIANGTVPGCPACKNVCD 1579  
 Db 1610 TGPILLGGVPLDPSFPVPMROFVGCNRLOVDSHIMADFIANGTVPGCPACKNVCD 1669  
 QY 1580 SNTCHANGGTCVQVQAFSCBCEPLGFGSKSCAOEMANPOHFLSSVVAWHGLSLPISQPY 1639  
 Db 1670 SNTCHANGGTCVQVQAFSCBCEPLGFGSKSCAOEMANPOHFLSSVVAWHGLSLPISQPY 1729  
 QY 1640 ISLMFRTRQADVLLQAITRGRSTITLQIREGVNLSVEGTGLQASSLRLEBRANDGDW 1699  
 Db 1730 ISLMFRTRQADVLLQAITRGRSTITLQIREGVNLSVEGTGLQASSLRLEBRANDGDW 1789  
 QY 1700 HHAQALASGSGR-----ALSPDYGOQARBNGLGRHMLHNSITVGGIIPBAGV 1754  
 Db 1790 HHAQALASGSGR-----ALSPDYGOQARBNGLGRHMLHNSITVGGIIPBAGV 1848  
 QY 1755 ARGFRGCLQGVSVSTPREGVNSLDPHSHESINVEGCSLPDPCDNPCCPANSYCSNDMS 1814  
 Db 1849 ARGFRGCLQGVSVSTPREGVNSLDPHSHESINVEGCSLPDPCDNPCCPANSYCSNDMS 1908  
 QY 1815 YSCSDPGYGDNCCTNVCDLNPCEHOSVCTRRSPAPHGTYCECPNYLAGPYCETRIIDPC 1874  
 Db 1909 YSCSDPGYGDNCCTNVCDLNPCEHOSVCTRRSPAPHGTYCECPNYLAGPYCETRIIDPC 1968  
 QY 1875 PRGMWGHPTGRCNCNDVSGRPPDCNKTSGBCHEKBNHRRPAGSP7CILDCYCPYGSLSR 1934  
 Db 1969 PRGMWGHPTGRCNCNDVSGRPPDCNKTSGBCHEKBNHRRPAGSP7CILDCYCPYGSLSR 2028  
 QY 1935 VCDPDEGQCPCKPBGVIGRCDCRCDNPPFAVTTNNGEVNYSCEPRA1EAGIWWPRTFGLP 1994  
 Db 2029 VCDPDEGQCPCKPBGVIGRCDCRCDNPPFAVTTNNGEVNYSCEPRA1EAGIWWPRTFGLP 2088  
 QY 1995 AAAPCKGSGFTAVRHCDEHGMPLPNL1FNTSITP5BLKFAERLQNNESGLDSGRSQ 2054  
 Db 2089 AAAPCKGSGFTAVRHCDEHGMPLPNL1FNTSITP5BLKFAERLQNNESGLDSGRSQ 2148  
 QY 2055 LALLLNATQHTAGVFGSDVKAAYOLATBL1HBS10RSGFGLSATODVHFTENL1LVGSA 2114  
 Db 2149 LALLLNATQHTAGVFGSDVKAAYOLATBL1HBS10RSGFGLSATODVHFTENL1LVGSA 2208  
 QY 2115 LLDFTANKRMBEL1QOTEGTAWLLOHYEAYASALQNNHHTYLSPTTIVPNI1VSVBL 2174  
 Db 2209 LLDFTANKRMBEL1QOTEGTAWLLOHYEAYASALQNNHHTYLSPTTIVPNI1VSVBL 2268  
 QY 2175 DKGNPAGALPRYEA1RGQPPDL1ETTVL1PESV-----R1TPPVVR-----PACP 2221

```

Db      2269  DKLNTGQVFPFBEOELPRELESSVFPADTFPPEKKGFPVRLNRRATPLTQAQ 2328
Qy      2222  GEAQPEELARQRHRELSQGBAASVITVTLAGLLPHNDPDKRSIAVRKPIITMP 2281
Db      2329  EPRARETSSSRKRRHPDEGQFPAALVITVTLQOLPEHDPHRSIRLRPNRVITMP 2388
Qy      2282  VVSIHVDDEBLPPALDKPVTVQFRLLETEREPTKICVFNMSILVSGTGMARGCEV 2341
Db      2389  VVSANVYSEGTFPLPSIORPLIVERFLETTERSKPVCVFNMSILDTGTGMSAKGCEL 2448
Qy      2342  VFRNSHYSCQCNMTSPVLMVDSRRNGEILPLKTLTYVALGVTALALLTFFPLTL 2401
Db      2449  LSRNTHVTCQCSHSAQVIMDIRRBRGEPLPKITVYALSLVALVAFVLSIV 2508
Qy      2402  RILRNOGIRNLTALGLAQVFLGNOADLPACVITAIIHLPLTCFSNALLA 2461
Db      2509  RILRNSHLSIKHKLIALPFSOLIFWVGINOTENPPLCTVAILHYVSMGTFATLVEN 2568
Qy      2462  LHLVRLTEVDVNTGPMKFTYMLGMPVAFITGLAVGLDEPGYGNPFCMLSIYDTLIM 2521
Db      2569  LHVYMLTEVRNIDGPMKFTYVVGMIPIAVITGLAVGLDPOGYGNPFCMLSIYDTLIM 2628
Qy      2522  SPAGVAAVAKSVLYTLAARASCAQROGFEKKGPVSGLOPSAVILLISATVLLAL 2581
Db      2629  SPAGVGVIIIVITVYLSAKVSCQRKHYYERKGVSMRTAFILLLVATVLLGL 2688
Qy      2582  SVNSDTLLFHYLPATCNCIOGPEIFLASYVLSKVRKALK-LACRKRSPDPAITTKST- 2639
Db      2699  AVNSDTLSFHYLPATCNCIOGPEIFLASYVLSKVRKALK-LACRKRSPDPAITTKST- 2748
Qy      2640  LTVSYNCPSPYADG--RLYOPYDGSAGSLHSTSRSGKSPSYIPLIREESALNGQGP 2697
Db      2749  LTVSYNCPSPYADG--RLYOPYDGSAGSLHSTSRSGKSPSYIPLIREESALNGQGP 2798
Qy      2698  GLGDPGSLFL-EGDQDQDPTDSDSLDSDOSGYASTHSDSEEEBEEBAAP 2756
Db      2799  NHGPDSTSFIRNSKKAHGPDSDSSELSL-DEHSSSYASSHTSSEDDGGEABDK 2853
Qy      2757  GEQGDMSLLGPGARLPLHSTPK-DGPGPGKAPMPGD--FGTTAKE----- 2800
Db      2854  WNPAGGA-----HSTPKADALANHYAPGPFDESLSGSDSEEDLDTPEHLKVETKV 2903
Qy      2801  SSGNGAPEERLRENGDALSRGSLGPLPGSSAOP--HKGLKKKCL--PTIS 2848
Db      2904  SVELHQOQGNHCGRPSPDPSGVLAQ-----PVAVLSQPOBQKGLKNKVTYPPPLP 2958
Qy      2849  EK--SLLRLPLEQGTSSRGS-SASESGSGGPPP-----RPPROSLQEOUNGWMP 2898
Db      2959  EQLKSRLEKLADEOSPTSSKRTSLGSDGVHATDCVITIKTPRPREGREHLNGV--- 3015
Qy      2899  AMSIKAGTVDEDSGSE 2915
Db      3016  AMNVRTGSAQANGSDSE 3032

```

RESULT 12  
US-10-150-811-70

```

; Sequence 70, Application US/10150811
; Publication No. US20040010120A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar et al.
; TITLE OF INVENTION: No. US20040010120A1 Polypeptides and Nucleic Acids Encoding Sam
; FILE REFERENCE: 15966-67CIPCON1
; CURRENT APPLICATION NUMBER: US/10/150,811
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 09/970,607
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/182,733
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/182,724
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/183,896

```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 60/184,497
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/224,157
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/184,482
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,744
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/197,083
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/233,405
; PRIOR FILING DATE: 2000-09-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-150-811-70

Query Match      57.7%; Score 8974.5; DB 12; Length 3034;
Beet local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

1 MRSPTAVPLPPTPPPLLLLLLLLLPPPLGDDVGPGRSGSRGSG-----S 48
Db      124  LRSRARGAEIRSP-----AVRSVPGGLDAL--CFPAAGGAAALSVLAETVFP 172
Qy      49  GACAPM-----GWLQ--PSSASNLMTYTSRCDATGTELTGHLVPHHDGLRVMCPSEA 99
Db      173  CSCCPVAGTCGRNPICLRPGSGAELALVCLGAAQA-----VWV----- 213
Qy      100  HPLPAPBEGCPWSCRLIGIGHLSPQKLTLPREHCLAPRLRCQCKLAQAPGLRAG 159
Db      214  -----ELVIOATSGTPBSPSV--SPSL-----LNLSP--RAG 243
Qy      160  --BRSPEESIGRRKXNVNTAPQGPSYQATVPENPACTFPVASTALIDPBEGARLE 217
Db      244  VRRS-----RGTGSSSTSPPLPSYQVSPENPAGTAVTELAHPDEDEGARLS 296
Qy      218  TYMDALFDSRNOFFSLDPTGAVTTAERLDRKSTHYVRVTVAQDHGMPRRSLATLTI 277
Db      297  YQWALFDERNSNGFLIDATGAVTTARSIDRETQTHVAKSAYVDGSPRRSAATYLT 356
Qy      278  LVYTDNDHVPFEQOETKESLRNLEVGIVLTVRATDGAAPPANAILYRLBSSGSSPS 337
Db      357  TVSPYNDHSPVFEQSEYRERIRRENLVGYEVLITRATDGDAPSNANRYYRLBAGAG 413
Qy      338  EVPEIDRSGVIRTRGVDEEVEBSVOLTYEASDQGRDPGRSTTAVFLSVEDNDNAP 397
Db      414  -VEIDARSGVAVTRAVVDEBAEYQLVLEANDQGNPPLSASATVHLVVEDENNYR 472
Qy      398  QFSEKRYVQVRBDVTPGAVLRTASDRDKGSNAVHVHSIMSGNARQGYLDAQTGALD 457
Db      473  QFSEKRYVQVRBDVAVMTAVLKVQAFDRDQGNAAIHYSIVSGNLKGQYLLHSLSSLD 532
Qy      458  VSPBLDYETTKETTLVRAQDGRPPISVNSGLVTVQVLDINDNAPIFVSTPFOATVLES 517
Db      533  VINPLDPEAREYTLIKAQDGRPPILNSSLGIVSVQVLDVNDNAPFVSSPQAAVLN 592
Qy      518  VPLPGVLTAVQALIDANAGNARLEYRLAGH-----DEPPTINNGTG 560
Db      593  VPLGHSVTLQAVDADAGENARLQRYLVDTASTTVGGSSVDSNPNASAPFPQIHNSSG 652
Qy      561  WISVAELDRBYDFYSPGEARDHGTPALTASASVTVLDVNDNPPFTQPEYTVRLN 620
Db      653  WITVCAELDRBYEHSFGEAVDHSPPMSSASVITVLDVNDNPPMTQVYELRLN 712
Qy      621  EDAAGTVSVTVAVVDRDAHSVTTYQTSQNTNRRSITSQSGGLVSLALPDYKLERQ 680
Db      713  EDAAGSVSVTLTARDRANSVTTYQLTGNTNRRPALSSQSGGLITLALPLDYKLERQ 772

```

Qy	681	YVLAVTASDGRQDTAQCIVVWVTANTANTRPVEFQSHYTVVNNBRPGETTUVLLSANDGD	740
Db	773	YVLAVTASDGRSTRHQAQVFINVIDANTRPFOSSHHTVSVSEDRPVCSTIATTSADD	832
Qy	741	TGENARITYFEDSIPOFRIDADTGAVTTOALEDEVDQSVYTLAITADNGIPKSDITY	800
Db	833	TGENARITYVEDPVPQFRIDPDNCTIYTMTELDYEDQAAVTIATADNGIPKSDITS	892
Qy	801	LEILVNDVNDMAPOFLRDSYQGSYVEDVPPFTSVQIATRDRDGLAKRVTYTTQGGDGG	860
Db	893	LEILILANDAPRLRDPYQGSVEDEAPSTSVQVATRDRDGFNRLLYTTQGGDGG	952
Qy	861	DGDFIVESTSGIVTTLRLDRDENVAQYLRAYAVDKGPR-DARPEMTYVVLVNDNPP	919
Db	953	DGDVIETSTSGVITQRDLDRENNAVYMLMALADRGSPNPLSASVGIQVSLVNDNPP	1012
Qy	920	VFEODEPDPVEBNSPIGLAVAVATYATDPDEGTNAQIWTQIVEGNIPEVFOULIFSGELT	979
Db	1013	VFEODELELFEVENSPIGSAVARIRANDPDGPNAGIITQIVEGNPEVFOULIDSGLR	1072
Qy	980	ALVOLDREDEPYLVLIQATSAPLYSRATYHVRLDRDNDPPVLCNEPILFNNTYTNSS	1039
Db	1073	ALVELDEVERDDYMLVVOGATSAPLSRATVHIRLLDQNDNPELPDPFILFNNTYTNKS	1132
Qy	1040	SFPGGALGRVAHPDIDSILTYSSEBANGELSLVYLNASTGEKLKSRALDNNRPLEATMS	1099
Db	1133	SFPGSGVIGRIYAHBDPLSDSLNTYTLQGENEISLLDPAITGELQJSRDLNNRPLEALME	1192
Qy	1100	VLVSDGVHVSATQACALRVITITDEMLTHSITLRLDMSPERFLSPLGLFOAVAAATLAT	1158
Db	1193	VSVSDGHSYATLCTLRVTIITDDMLTNSITVRLNMSQEKRLSPLSLFVBGAVATVLT	1252
Qy	1160	PPDHVVVNVQBDPAFBGHILNLSLSVGQPPRPGGPPPLPBSBLQERLITANSILTAI	1219
Db	1253	TKDQIIPVFNIQNDYV-SSNITLMTFSAALRPGTRG--RFPESBDLOQIYANRLTLTI	1309
Qy	1220	SAQRVLPPDDNI CLREPCENMRCVSVLRPFSSAPFIASSSLPFIHPVGRLRCRCPG	1279
Db	1310	SAQRVLPPDDNI CLREPCENMKCVSVLRPFSSAPFISSTVLPRIHPITGLRCRCPG	1365
Qy	1280	FTGDYCEYEDVLCSYRPGCPGHRCKSRBGGYTLCLRDGYTGEBHCEVSARSGRCTPGVCK	1339
Db	1370	FTGDYCEYEDVLCYSPNPGANGRCRSRGGYTCBCEPDFTGBHCQVNVRSGRCASGVCK	1429
Qy	1340	GGTGVNLLVGFKCDCESDPEKPRCOYTTSSFPKHSITTRGLRQRFHFLTALSARK	1399
Db	1430	GGTGVNLLVGFKCDCESDPEKPRCEVSTSSFPQSVTRGLRQRFHFLTALSARQD	1489
Qy	1400	RDGLLLYNGRFBKXDPVALFVIOQOVLTFSAGSITTVSPFVGVGSDOQMTVOLY	1459
Db	1490	RMALLIYNGRFBKXDPVALFVIOQOVLTFSAGSITTVTPQVPGVSDGRKMSVLQY	1549
Qy	1460	YKNPILLGQYGLPQGSSEQKVAVVTYDGGDTGVALRFGSVLGNYSCAAQGTQGSKSLDL	1519
Db	1550	YKNPENIGHILGPHGSGSEKVAVVTYDCCDAVAVHFGSVYGNYSCAAQGTQGSKSLDL	1609
Qy	1520	TGPIILLGVPRLPBEPFYRMRQFVGCMMNLQVDSHIDMADPIANNGVPRCPAKKNCV	1579
Db	1610	TGPIILLGVPRLPBEPFYVHSHQFVGCMMNLSDIGITVMAAFIANGRRACASGRNCPD	1669
Qy	1580	SNTGANGGTCVNOMDAFSCBCEPLGFGKGSCAQEMANPOHFIQSSILVAHGLSLPISQPMY	1639
Db	1670	GTSQNGGTCVNRMWTVCEBCEPLRAGKNCQDAMHPRPRTGBESVYVMSDDITTSVMY	1729
Qy	1640	LSLMFRTQADGVLLQAITRGRSTITLQLRBGMVLSVEGTGLQASSLRLEBRANDGM	1699
Db	1730	LGLMFRTKEBGMLEATAGTSSRLHQLIINSYIRFEVSYSPSDVASVQSKSRITDQGM	1789
Qy	1700	HHQALALGASGGPRH-----ALSPDYQQAQABNGMLGRLLGHLHNSITVGGIPEBPAGV	1754
Db	1790	HHLLIEL-RSAKSGKODIYTLAMTIDYMDQDSVQIQLGQLPGLKRRITVIGGVEDKVS	1848

QY	1755	ARGRGGLQGVARSVDTEBGNVSLDPSHSGBSINVEQGSLSLDPDCSNCSNPNVSYNSNMD	1814
Db	1849	RHGRGCMQGVNRGERSITNATITAMNDALKVRYVQDCDVEDPCLASSCCPPHRRPCRDYWD	1908
QY	1815	YSCSGCDPGYTGDNCTNVCYNCPNCEHOSVCTRKESAPRHYTCBCPNYLAGPYCETRIDOPC	1874
Db	1909	YSCTCDHGYGKKCVDACLNPKCHVAACVRSPTPTPGYSCCGPHGYGQCEKKVLLPC	1968
QY	1875	PRGWMHPTQGPCNCDVSKSPDPDNCNTSGCKENHYPBGSFCLLDCDCTYGSLSR	1934
Db	1969	PKGMWGNPVCGPCAACVSGQDPDPCNTNQCCQCKENYTPRPADACLPCDCPFHGSISR	2028
QY	1935	VCDPEDDQCECKPGLVLRQCDRCNDPFAKYTNGCEVNVNCSPRALBAGIWMPPTRFGLP	1994
Db	2029	ACMDYDQCACKKPGVILQRCNRCNDPFAKYTNGCEVNVNCSPRALBAGIWMPPTRFGLP	2088
QY	1995	AAAPCPYSGTGTAVRHDDEHGMPLPPLFNCTSTSTSEBKGFAPRLQPNESGLDGSRSQ	2054
Db	2089	AAVPCPGPSGNAVRHCSGKERGMPLPPLFNCTSGSFVDLVAWEKLNRMNTRMDGNRSLR	2148
QY	2055	LALLRNATHTAGYFGSDVKNVAYQLATRLLAHSTSGRFGSLSTODVHPFNMLTRVGS	2114
Db	2149	LAKLRNATQGNSTLPGNDRTAYQLARILQHSRSQGDPLATREANFHDVHTGSA	2208
QY	2115	LLDYANRHMELIQOTEGGTAMLLQHYEAYASALQNMRTYLSPTTYENIVISVRL	2174
Db	2209	LAPATASWEQIORSBAGAAQLRHEAFYSNVARVKTYLRPFYIYATNMIAVDIF	2268
QY	2175	DKGNPAAKLPYREALRGEQPPOLETIVILPESV-----RTPPVNR-----PAGP	2221
Db	2269	DKLNPFTAQVPRFEDIOEBLPRLEBSVSPADTFKPEKKEGVVRLTNRRTPLTAQP	2328
QY	2222	GEAOPPEELRQRORHRELQGEAVVAVIYRTLAGLLPHNYDPDKSLRVPKPIINTP	2281
Db	2329	EPRAERETSSRRRRHDBEQFAVALVYITRLGQLPHEYDHPDHSRLRPNRPVINTP	2388
QY	2282	VVSISVHDEBELPRALDKVYVYQFLLRFBERTKPICVWNHNSILVSGTSGMSARCEV	2341
Db	2389	VVSAMVYSBETPPESSIQRPILVEFSLLETBERRKPCVPMNHSLLDVGTTGSGMSAKCEL	2448
QY	2342	VFRHNSNSCQCNHMTSPATLVMDVRSRENEBILPCLKTYVAAGVTALALTFEFTLL	2401
Db	2449	LSRRRTHTVQCSHMSAVLMDISRREHREVLPKTIYTAALSLVALLVAVFVLSV	2508
QY	2402	RIIRSNQHGIRRNITAAQLAQLVFLGIGNADLPFACVYAIILHFLYLCFSWALLBA	2461
Db	2509	RIIRSNHSHHKMLIALPFSQILFMVGINQTEPFCVVALILHYSWMTGFAMTIVEN	2568
QY	2462	LHLVRALTEYRDVNTGVMRFYVMLGMQVPAFETGLANGLPBEGGNDFCMLSXYDTLIW	2521
Db	2569	LHLVRMLTEYRNIDTGMRFYHVGWQIPALVGLAAGLPDQGNDFCMLSXYDTLIW	2628
QY	2522	SFAPGPVAFVMSVFLYTLAARASCQAORQGFKEKGPVSGLOPFAVLLLSATMTALAL	2581
Db	2629	SFAPGPVGTIILNIVIPVLSAKVSCQKHHYERKGVVSMLTRFALLLLVYATMTLGLL	2688
QY	2582	SVNSDITLFRYLPATCNCIGPPIFLSYVLSKVRALK-LACSRKPSDPALTTYST-	2639
Db	2689	AVNSDITLSFYLPFAFSCIQGIFVLLPHCAHRRVRHMLRAVLAKGLQULDSAYTTATL	2748
QY	2640	LTSYNSCPSPADP--LTYQPYGSAGLSTSGSGQSPYIPFLRREBSALNPGQGP	2697
Db	2749	LTRILNCTNMTYSBEPDLKTRALGESTASLDSTTRDEVO-----KLSVSGGARG	2798
QY	2698	GLGDPGSLFL-EGDDQOHDPDTDSLDLSLBDQSGSYASTHSSDSREBEBEBEBAFP	2756
Db	2799	NHGEPDTSFIPRNSKKAHGPDSDSELSL-DEHSSSYASHTSDSDDEGAEADK----	2853
QY	2757	GEQGMDSLILPGARLPLHSTPK--DGAPGQKAPWPD--FGTAK-----	2800
Db	2854	-----WNPAGCPA-----HSTPKADALANVPACMPBESLAGSDSEBLDTRPHLKVETKY	2903
QY	2801	-----SGNGAPBEERLRENQALSESGSLGPPGSSAOP-----HKGILKKCL--PLIS	2848

Db 2904 SVELHROQVHNCGRBPSDESGVLAK-----FVALVSQPOEQKGLIKKVVTPPLP 2958  
Qy 2849 EK--SSLRLPLEOCTGSSRGSS-SASBGRGPP-----RPPROSLQOLGVMP 2898  
Db 2959 EOPLRKREKLADQSPSTSRSSLSGSDGVNATDVTITKTBREREGREHLNGV--- 3015  
Qy 2899 AMSIKATVDEDSGSGE 2915  
Db 3016 AMNVVTGSAQANGSDSE 3032  
  
RESULT 13  
US-09-737-149-2  
; Sequence 2, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Spyrek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,544  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 2000-01-04  
; PRIOR APPLICATION NUMBER: 60/174,962  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 3014  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-737-149-2  
  
Query Match 56.3%; Score 8754; DB 9; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  
  
Qy 12 TPRPRLILLLILPRLIGDOVGPGRSIGSRGSSGACAMKLCPSASNLMLYTSR 71  
Db 57 TPRARRELL-----DVGRDGRLAGR-REVSGAGRP-----PLOYVLY----- 93  
Qy 72 CADAGTEILGNLVRHNDLGRVWCSEBAHPIPLRPAEGCWSGRLIGGHLSPQKLT 131  
Db 94 ARSATATLSRRL-----RAKTLR-----GGARARLCGTGARLC--GALCF 133  
Qy 132 P-----EENPCLKA-----PRLRCOS-----CKLAQARG- 155  
Db 134 PVGGCAAAQHGALAPRTLLPACRCRPRRRCRPGPICLPFGSGVRLRLCALARRA 193  
Qy 156 LKAG-----ERSPESLGGRRKRVNTLP-----QFPPSYQAT 189  
Db 194 VAVGALAEATAGTSPASPSRPLRPMLEARARPARARRGTSGRSLKPMENYQVA 253  
Qy 190 VENOPAGTPVASTLAIDPDEGAGRLLEYTMDALPDSRNOFPSSLDPTVGAATTA 249  
Db 254 LPENPAGTLILQILAHYTIIBEEERAVSYTMEGLPDESRGYFRIDSATGAIST 313  
Qy 250 ETKSTHVRVTAQDHGMPRRSALATLTLLVTDTNHDVFEQOEYKESLRENTLE 309

Db 314 ETKETHLRVAADVSYSPPSASATYITVLYVDITNDHSPVFEQSRERKRVENLE 373  
Qy 310 TVRATDGPAPNANILYRLLEGSGSPSEVEIDPSRGVIRTRGPVDRBEVSYQ 369  
Db 374 TIRASDRSPINANILRYVLGAW-----DVQLNBSGGVSTRAVLDRBEAAVE 429  
Qy 370 SDOGRDGPSTTAAVPLSVEDDNDNAPOSEKRYVYQVNEQDYTPGAPVLA 429  
Db 430 NDGRNPGPLSATVYVIEVEDENDNYPQSEBQNVVQVQVDEGLNATVLEQ 489  
Qy 430 SNAVHTSIMGARQGFYLDACQALDVVSPLDYETTKETKYLAVRAODGGR 489  
Db 490 QNAIHYSILSGNAGGYLHSLGSLIDVINEPDEVOGKYSLSIKQDGRPPLINS 549  
Qy 490 LVTYQVLDINDNADIVSTPPQATVLESVPGLVLTIVQALDADGNARLE 545  
Db 550 VVSQVQVLDVNDNEPIFVSSPFOATVLENVPLGYPVVHIOAVDADSGENAR 609  
Qy 546 ---GVGH-----DPEPTINNGWTISVAELDRBEVDPSFGVEARDHCT 592  
Db 610 TFLGGSSAGPPNAPPTDPFPQIHNSSGMWTVCAELDRBEVHTSFGVEARDH 669  
Qy 593 SASVSVTVLDVNDNPTFTQPEYTVRLNEDAAVSTSVVTSADVADASVIT 652  
Db 670 STSVITVLDVNDNPTFTQPEYTVRLNEDAAVSTSVVTSADVADASVIT 729  
Qy 653 RNRFTSISQSGGGLVSLALPLDYLERQYVLAFTASGTRQDTAQIVNVTANT 712  
Db 730 RNRFLSSORGGGLITLALPLDYERQYVLAFTASGTRQDTAQIVNVTANT 789  
Qy 713 QSSHVTYVNDNDRPAGTIVLISATDDEGENARITFPMBSITPOPIIDVGT 772  
Db 790 QSHRTVSVSDRKGVSITATLSANDDEGENARITVIOVPQFIDPDSGMYT 849  
Qy 773 LDYEDQVSYTLATARDNGIPQKSDITYLRLVNDVNDNAPOFLRDSYQSG 832  
Db 850 LDYENQVAYTLTIAQNGIPQKSDITTLRLILDANDNAPQFLMDPYQSG 909  
Qy 833 SVLQISATDSDSGANGVFTTQSGDGDGDFYESTSGIVRTLRILDRVNAQ 892  
Db 910 SILQVSATDRDSGNGLLYTFQSGDGDGDFYESTSGIVRTLRILDRVNAQ 969  
Qy 893 AVDRGMP-PARTPEVTVTVLDVNDNPPVFEODEFDFVENSISGLAVARVTA 951  
Db 970 AVDRGSPPLASAVETQVTLIDINDNAPMEKQLELFEENNVGSVAKIRAND 1029  
Qy 952 TNAQIMQIVEGNIPVFOLDIPSGELTALVDLYEDRPEVTVLOATSAPLV 1011  
Db 1030 PNAQIMQIVEGDMRHFFQDLNLNGDLRAVVELDFVRAREVTVLVQATSA 1089  
Qy 1012 RLDRNNDPPLNAGFELLFNNTYNNRSSPFGAIGVPAHPADPISLTSY 1071  
Db 1090 LLVQNDNPPVLPFOILFNNTYNNKNSPPTGYIGCPADHPVDSILNTY 1149  
Qy 1072 LVLNASTGELKLSRALDNNRPLEAIVSVSDVSHSVTQACALRYITIDEM 1131  
Db 1150 LLLNDPATGELQLSRDLNDRPLEALMEVSDSHSVTAFCTIRVITIIDM 1209  
Qy 1132 RLBDMSPERLSPLGLFIQAVATLTATPDHVVVFNVOBTDAPGGHILN 1191  
Db 1210 RLBNMSQEKFLSPILALFVGVAAVLTSTKDDVVFVNVQNDTDV-- 1268  
Qy 1192 GPGGPPPLPSBEDQELVYLNRSLLTAISAOVLPFPDNI CLAREPCEN 1251  
Db 1269 GVRG--QFPSEEDQEOEIVYLNRTLLITISORVLPFPDNI CLAREPC 1336  
Qy 1252 SAPFIASSSVLFRPIHVGGLRCGCPGFGDYCEYVDCYSRPPGHCRCR 1311  
Db 1327 SAPFLSSTVLFRIHINGLRCGCPGFGDYCEYVDCYSRPPGHCRCR 1386  
Qy 1312 CLCRDGYTGEHCVSABSGCTPGVCNNGGTCVNLLVGFGRKDCPSGD 1371



Db 1387 CECFEDFTEGHECEVDASGRGANGCKNGGTGVNLLIGFPHCVCPPEGEYRPYCEVTTSS 1446  
Qy 1372 PPAHSPTFRGLRQRFHFTLALSPATKERDGLLYNGRFEKHDPALEVIQOVLTS 1431  
Db 1447 PPOQSIVTRGLRQRFHFTLALSPATKERDGLLYNGRFEKHDPALEVIQOVLTS 1506  
Qy 1432 AGESITTSBPVGVSDQGMHTVOLKYNNKPLLQGTGLPOGSEKQVAVVTVGDCDTGV 1491  
Db 1507 AGETTTTVAAPKVSVDGRMHSVQYVYKPNIGHGLPHGSGSKMAVVTVDDCTTM 1566  
Qy 1492 AARSGVLNYSGLAAGTQGSKSLDLTGPILLGVVPLPESFPYRMQOFCQNRLOY 1551  
Db 1567 AAFEGDIGNYSGLAAGTQGSKSLDLTGPILLGVVPLPESFPYRMQOFCQNRLOY 1626  
Qy 1552 DSRHIMADFIANGTVPGCPAKGVCDSTNGAGTCVQOMDFAFGCEGLGFGKSCAO 1611  
Db 1627 DGNVDMAGIANGTVPGCPAKGVCDSTNGAGTCVQOMDFAFGCEGLGFGKSCAO 1686  
Qy 1612 EMANPQHLGSSLVANHGLSLPISQPMYLSLMFRTQADGVLLQAITRGRSTITLQLRBG 1671  
Db 1687 AMPHPQLFSGESVYMSDLNIIISVPMYGLMFRTRKEDSVLMEATSGGPTSRLLQILNN 1746  
Qy 1672 HVMVLESTGLQASSRLBERGANDGMHQAOL--GASGPRGAL-LSFPGQQRAB 1727  
Db 1747 YLQFEVSHGSDVESVWLSGLRVTDEMHLLIETLNKXEDSEMKHLYMTTLDGMDQNK 1806  
Qy 1728 GNLGPRHLGLHLNITVGGIRGPAAGVARGFCLOGVRSYDTPREGVNSLDPSHGSIYV 1787  
Db 1807 ADIGMLPGLTVASVYVVGASBDRKSVRRFGRCQMGVNGGTFVTVAITLMMNNALKAVY 1866  
Qy 1788 BQGSLLPDCDSNCPANSYCSNDMDYSKSCDPGYGDNCTVCDLNPCEHGSVCTKRP 1847  
Db 1867 KQCDVDVDDPCTSPCPNSRCHDAMEDYSVCCKGTLGNCVDAQHLNCPENNCAVCVRSP 1926  
Qy 1848 SAHGYTCBPBPVNYLBYCETRIDOCBPGRMGHPFCGPNCEVSGKGPBDCNKTSGBE 1907  
Db 1927 GSHQGVYCEGSHYBPCKNKLDLPCPRGMGNPVCSPGCHAVSKGFPBDCNKTKNGQCO 1986  
Qy 1908 CKNHRYRPPGSPYCLLCDYPTGSLSRVCDPEPDQCEPKKPGVIGRCCDRCNDPFAVTTN 1967  
Db 1987 CKENYKTLAODTCLPCDCPFGHSHRTCDMATGQCAKRGVIGRCCNCRNDPFAVTTN 2046  
Qy 1968 GCEVNTDSCPRALBAGIWMPTRPGLPAAAPCEKSGFATAVRHCDERHGLPMLFNCTG 2027  
Db 2047 GCEVINYGCPRKAEAGIWMPTKFGQPAVPCPKSGVANAVRHCDERHGLPMLFNCTG 2106  
Qy 2028 ITSEELKGFARERQNRSGLSGRSQALLANATQHTAGYGSVYKAYOATLTLAH 2087  
Db 2107 ISFVDLRANNEKLSRNETOVDGARALQVRLASATQHTGTLFGNDVTRAYQLGHVLR 2166  
Qy 2088 ESTQRFGLSATQDVHTENILRVGSLADTANKRHMLIQTEGGTAMLLQHYEAYASA 2147  
Db 2167 ESNQGFGLDAAITQDAPHEDEVIIHSGSALLAPATRAMEOQREBEGTQOLRLBEYFSN 2226  
Qy 2148 LAONMHTVYSPPTVTVNIVISVRLDKGNPAGAKIPRYEALRGSEOPDLFTTVILPES 2207  
Db 2227 VAAENVRTYLRPVIVYANMILAVDIFDKENFGARVPRDTHIEEPRELESSVFPAD 2286  
Qy 2208 VFR-----ETPVVRPPG-----PGEAQBPEELARQRHRRHDELSQGEAVASVIT 2254  
Db 2287 FFRPREKEGRLRPPGRRRTTTPOTTRPGPOTERBAPISRRRRRPPDDGQPAVALVIT 2346  
Qy 2255 IAGLLPHNYDPRKSLRVPKRPINTPEVVISVHDEBELPRALDXPVTVQFRLLETER 2314  
Db 2347 LGQLLPRYDPRKSLRVPKRPINTPEVVISVHDEBELPRALDXPVTVQFRLLETER 2406  
Qy 2315 TKRIVCFPMNHSILVSGTGWSSARGCEVFPNESHVSQCCNMTSPAVLNDVSRREGE 2374  
Db 2407 TKRIVCFPMNHSILVSGTGWSSARGCEVFPNESHVSQCCNMTSPAVLNDVSRREGE 2466  
Qy 2375 PLKTLTVVAGVTLAALLTFPFLTLRLIRSNQHGIRMLTAALGLAOLVPLGINQAD 2434  
Db 2467 PLKTLTVVAGVTLAALLTFPFLTLRLIRSNQHGIRMLTAALGLAOLVPLGINQAD 2526

Qy 2435 LPFACVTAIHLHFLYLCFESWALLLEALHYRALTEVDVNTGPMRFYTMGMGVPAFT 2494  
Db 2527 NFPLTVVAILHLYHMTFTWMTLVESLHVRLMTRKXNDIDGPRFTYVGMGIPALVT 2586  
Qy 2495 GLAVGLDPEGYNPDCWLSYDYLWSPAGVAPAVMSVFLYLIAARASCAORQFBE 2554  
Db 2587 GLAVGLDPEGYNPDCWLSYDYLWSPAGVAPAVMSVFLYLIAARASCAORQFBE 2646  
Qy 2555 KKGVSGLQSPAVYLLLSATWLLALSVDNLTLLFHLFATCNCIQPFTLSVYLSK 2614  
Db 2647 KKGVSGLQSPAVYLLLSATWLLALSVDNLTLLFHLFATCNCIQPFTLSVYLSK 2706  
Qy 2615 EYRKALK-LACSRKSPDPALTTKST-LTSSYNCPSPADG--RLYQFPGDSAGSLHSTS 2670  
Db 2707 EYRKALKVLAGRKALHEDSATTRATTLTRSLNCTMTGDEGDMKRTLGSTABLDSTIV 2766  
Qy 2671 RS-----GKSQPSYIPPLLREBSALNPGQPPGLGDPGLFLEGDDQ 2713  
Db 2767 RDEGIQKLGVSGLVRSNGHGEPPDASLMPRSCD-----PPG----- 2802  
Qy 2714 HDPDSDSDSLDSDOGSVASTSSDSBEBEBEBEBEAAAPRGQMDSLGPAERLP 2773  
Db 2803 --HDSDSSELSL--DEQSSSYASHSSDSEDDGVABEK-----WDPARG-----A 2845  
Qy 2774 LHSTPK-DGCPQPKAPWPGDFTAKESGNGAP-----BEERLRENGDAL 2818  
Db 2846 VASTPKGAVANHVAPGPDGSLASDSEDPGKRLKVTYKVSYLEHREBQSGHRYVP 2905  
Qy 2819 SRESGLGPLSSAQP--HGILKKCL--PTISKS--SLRLPLEQCTGSSRGS 2869  
Db 2906 PQESGGAARLASQPPQKGIILNKRTYPPPLTLTEQTLKGRREKLADCEQSPTSR 2965  
Qy 2870 ASESGRGP---PPRPPROSLQELNGWPMIANSIAGTVDESSGE 2915  
Db 2966 TSLSGGPDCAITYKSPEREGRDLNGV---ANAVRTGSAQADGSDSE 3012

## RESULT 14

US-10-241-220-107  
; Sequence 107, Application US/10241220  
; Publication No. US20030148408A1  
; GENERAL INFORMATION:  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Phillips, Heidi  
; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF TUMOR  
; FILE REFERENCE: P5010R1-US  
; CURRENT APPLICATION NUMBER: US/10/241,220  
; CURRENT FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 120  
; SEQ ID NO 107  
; LENGTH: 3014  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-241-220-107

Query Match 56.3%; Score 8754; DB 12; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;

Qy 12 TPEPPLILILILILPPLIGQVGPGRSLGSRGSSGACADMGWLCPBSSANMLYTR 71  
Db 57 TPRAPRELL-----DVGRDGLRAGR--RVSAGAGRL-----PLQVRLV----- 93  
Qy 72 CNDAGTELGHVPHHNDGLRVWCPRSESAHILPRLPABCCPMSCRLGTGCHLSPPGKLT 131

Db 94 ARSAPATLSRL-----RARTHP-----GCGARALCGTARLC--GALCF 133  
 Qy 132 P-----BEHPCLAK-----PRLRCQS-----CMAQAPG- 155  
 Db 134 PVBGGCAAAQHSLAAPTLLPACRCPRRPRRCRGPICLPBGSGVRLMLCLARRBAABA 193  
 Qy 156 LRAG-----ERSPBESLGGRRKRVNTAP---QOPPSYQAT 189  
 Db 194 VAVGALAEATAGTPASBPSPPLPPNLPBARAGARARARCTSGRGLCKPMPMYQVA 253  
 Qy 190 VBNOPAGTPVASIRAIIDPDEGAGRELEYTMDALPFRSBNQPSLDPVYGAATTAELRLR 249  
 Db 254 LEBENBAGTLIIQAHAYHTIEGSEBVSSTYMEGLFEBERSRGIFRISATGAVSTDSVILR 313  
 Qy 250 ETKSTHVFVTAQDHGMPRRSALATLLIIVTDNTNHD PVFBOOEYKESIRENLEVEYEL 309  
 Db 314 ETEETHVLKAVADVSTPPRSATTTIVLYKOTNDS PVFBOGSEYERERENLEVEYEL 373  
 Qy 310 TYRAITDGDAPPVNANILYRLBESGSPSVFELIDPRSGVIRTRGPVDRREBVSYOLTVEA 369  
 Db 374 TIRASRDRBPINANIRYRLVGAM---DVFOINBESSGVASTRAVLDRERAAEYQLVBA 429  
 Qy 370 SPOGRDPPGRSTTAAVFLSVEDNDNNAPOPSERKYVVOYREBDVTPGAPVLTATASDRDG 429  
 Db 430 NDOGRNPGLSATATVYIEVEDENDNYPOPSQNTVVOYVEDGANTAVLRVQATDRDG 489  
 Qy 430 SNAVVHYSIMSGNARQOFLDAGTALDVSPPLDYETTEKYLTRYAODGGRPLSNVSG 489  
 Db 490 QNAALHYSITLSCGNVAGQFLHSLSGILDVINPLDPEDVOKYSLISIAQOGGRPLINSRG 549  
 Qy 490 LVTVOVLIDNDNAPIFVSTFPOATVLESYPLGLVLAHVOALIDAGDNARLEYRLA---- 545  
 Db 550 VVSVOVLVDNDNEPIFVSSPFOATVLENVPLGYPVHIOADVADSENNARLEYRLVDTAS 609  
 Qy 546 ---GVGH-----DEPTINNGTGMISVAABEDREKVD FYSGVEARHDGTALTA 592  
 Db 610 TILGGSGAGPKPAPRPPDPFOIHSSSGHITVCAELDRKEVBHISGVEAVDHGSPMS 669  
 Qy 593 SASVSVTVLVDNDNPTPTQPEYVALNDAAVGTSVTVASAVDRDAHSVITYOITSGNT 652  
 Db 670 STSVSITVLVDNDNPVFTOPTYELRLNEDAAVGSVLTLOARDNDANSVITYOITSGNT 729  
 Qy 653 RRPESITSGGGGLVSLAPLDYKLEKOVVLAVTASDGTRODTAOTVANNVTANTHRPVF 712  
 Db 730 RRRFALSSORGGGLTTLALPLDYKOEQVYLAVTASDGTSHTAIVLINVDTANTHRPVF 789  
 Qy 713 QSSHTVANNEDRPAGTIVVLSATDEBGENARITTYPMEDSIPOFRIDADGATTOAE 772  
 Db 790 QSSHITTVSSEDRPVGTSTIATISANDEDEGENARITTYIQDEVPOFRIDPDGTYMTME 849  
 Qy 773 LDYEDQVSYTLAITARNDGIPOKSDTTYLEILVNDVNDNAPOFLRDSYOGSVYEDVPPT 832  
 Db 850 LDYENQVAVTTLTIMADNGIPOKSDTTYLEILANDNAPOFLMDPFGOSIFEDAPST 909  
 Qy 833 SVLQISANDROBGLNGRPVYTFQGGDDGDSTIVSTSGIVATRLDRBENVAQVYLAAY 892  
 Db 910 SILQVSATDRDSGPNRGLLYTFQGGDDGDGYIAPTSGVIRTORRLRENAVAVNLVAL 969  
 Qy 893 AVDKGMP-PARTPMEVTVTLVDNDNPPVFEODEBDFVYEENSPIGLAVARTATDPBG 951  
 Db 970 AVDRSPFPLASVSIQVITIIDINDNAPEKDELEFVEENNPVGSVYAKIRANDPBG 1029  
 Qy 952 TNAQIMYOIVEGNIBEVFOLDIFSGELTALVLDYEDRPEYVLTIOATSAPLVSRATVHV 1011  
 Db 1030 PNAQIMYOIVEGDMHFOLDILNDGLRAMVELDEPREVEYLVVQATSAPLVSATVHI 1089  
 Qy 1012 RLDDNDNDPVYAGRELLPNNVYTRSSSFPFGALIGRPYANDPDISGLTTSFREGNLS 1071  
 Db 1090 LLDVNDNDNPVLPDQILFNNVYTKNSFPVGVGCIPAHDPDVSISANTYTFVGNBLR 1149  
 Qy 1072 LVTLNASTGELKLSALDNNRPLKALMSVYLVSDGHSVTAOCALVNTIITBMLTHSITL 1131  
 Db 1150 LLLDLPATGELQSLSDNDNNRPLKALMSVSDGHSHTVAFCTLANVTIITDMLTNSITV 1209

Qy 1132 RLEDSPERFLSPLGLFLIOAAATLATPPDHVVFNVOJDADAGHILANLSVGOPE 1191  
 Db 1210 RLENNOSKELTSLPALTFVBGAVALSTTKDQVFPFNQNTDV--SNILANTFSLBGB 1268  
 Qy 1192 GPGGPPPLPSDLOERLYNRSLLTALSAORVLPDDNI CLRPCEYMCVSVLRPDS 1251  
 Db 1269 GVRG--QPPSEEDLOEQIYLRNTLLTISTQVLPFDNI CLRPCEYMCVSVLRPDS 1326  
 Qy 1252 SAPFLASSVLPRIHPVGGARCRCPGPTDGCETVLDICSRPCGPHRGRSREGGT 1311  
 Db 1327 SAPFLSSTVLPRIHPINGLRCRCPGPTDGCETVLDICVSDPGANGRGRSREGGT 1386  
 Qy 1312 CLCRDGYTGEHCEVARSARSGCTPVCCKNGTCVNLVAGFKDCDSSGFEKPYCQVTRS 1371  
 Db 1387 CECFEDPFGHECEVDARSGRANGVCCKNGTCVNLIGCFHCVCPPGEYERYCQVTRS 1446  
 Qy 1447 FPPQSFVTFRGLRQRFHTLITLTFATQBRNGLLYNGRFNEKHDFLALIEYDEQVLTFS 1506  
 Db 1432 AGESTTVSPVPGVSDGOMHTVQKYNKPILGQGLPOGSRQKXAVVTVDCDTCV 1491  
 Qy 1507 AGEITTVAPKVPVSGVSDGRHNSVQVQYKNPNIGHGLPHGSGEKXAVVTVDDCDDTM 1566  
 Db 1492 ALRFGSVLGNTSCAAQGTQGSKSLDLTGPLLGGVPLDPESFVVRORQFVCKRNTQV 1551  
 Qy 1567 AVRFKQDIGNVSCAAQGTQGSKSLDTGPLLGGVNLBDFPVHNRQFVCKRNTSV 1626  
 Db 1552 DSRHIDMADFTIANNGTVPGCPAKXNVCDSNTGNGTGVNQMDAFSCCPLGFGKSCAQ 1611  
 Qy 1627 DGKNVMAGFIANNGTREGCAARNFCDGRRCQNGTGVNMMNYLCEPLRFGKNCQ 1686  
 Db 1612 EMANPOHFLGSSIVAMHGLSLPISOPWTLSPMPTROADGVLQAITRGRSTIQLREG 1671  
 Qy 1687 AMPHOLEFSGSVSWSDDLNTIISPVTLGLMPTRKEDSVLMKETSGLPYSFRQIILNN 1746  
 Db 1672 HVMLSVEGTGLQASSLRLEPERANDGWHAQLAL--GASGGRGHAL--LSFDYGOQRAE 1727  
 Qy 1747 YLQFESHGPDVDSVMSGLRVTDGEMHLLIELKNVKEBSMKHLVMTLTDVGMQNK 1806  
 Db 1728 GNLGPRHLGHLNSITVGGIIPGAPGVARGRGLQGYRVSDTDEGVNSLDPHGESINV 1787  
 Qy 1807 ADIGMPLGLTVRSVVVVGASBDKVSVRKRGFGCMQGRMGCTPTNVAITLMNNALKRV 1866  
 Db 1768 EOGSLPDPDCNSNCPANSVCSNDMDYSGSCDPPGYGDNCTVNCDDLPCHGOSVCTKRP 1847  
 Qy 1867 KQGCDDVDDPCTSSPCFPNSRCHDMEDISCVCDKGYLGINCVDACHLNPCEMGA CVRSP 1926  
 Db 1848 SAPHYTCECPNPIAGPYCETRIDQPCPRGMMGHPTGCPNCNDVSKGDPDCNKTSGECH 1907  
 Qy 1927 GSPQGYCEGSPSHYGYCEMKKDLPCPRGMMGAPVGGPCHCANVSKGDPDCNKTNGCQ 1966  
 Db 1908 CKENHYRPGSPPTCLDCTYTGSLSRVCDPBDQCPCKKPVYIRQCDRCNPNPASTTN 1967  
 Qy 1987 CKENYKYLADOTCLPCDCPFGHSHRTCDMATQCAKCPVYIRQCDRCNPNPASTTN 2046  
 Db 1968 GCRYNVDSCPRAIAGIMPPTRFGLPAAAPCPGSGSTGAYRHCDERHGMPLPMLFNCTS 2027  
 Qy 2047 GCERYTVNGCPRAFGAGIMPPOTKFGQPAVPCPGSGISGNAVRHDSGKGMPLPMLFNCTT 2106  
 Db 2028 ITFSELGFAERLQRNESGLDSGRSQJLALLRNATQHTAGYFGSDVYKVAQOLATRLAH 2087  
 Qy 2107 ISFDLARNMEKLSRNETQYDGAVALQVVALRSATQHTGTLFNDVRYATQOLGHLVLOH 2166  
 Db 2088 ESTORGGLSATQDVHFTENLIRVGSALLDTANGHEHLLQOTGSGTAMLLQHTBAASA 2147  
 Qy 2167 ESWQGGFDLATOADHAEVYHSGSALLAPATTAABEQIORSBGTAQLRLRLEGYFSN 2226  
 Db 2148 LAQNNRHTVYSPFITVPNIVISVVRLDKGNFAGAKPRREALRGEOPDLETTVILPES 2207  
 Qy 2227 VARVRRTYLRFPVITYANMILAVDIFDKFNFTGARPRPDTIHBEFRRLBESSVSFPAD 2286

2208 VFR---ETPPVVRPAG-----PGEAOBEPEELARQRHPELSOGEBANASVIIYRT 2254  
2287 FFRPEPEKEGPIIRPAGRRATTPTQTRPGPOTEREARISRERRRDPDAGAPALVITIRT 2346  
2255 LAGLTHANYPDRGRSLVRPKRPIINTPVVISVHDBELLPRALDPEVTVQPFLLTEER 2314  
2347 LQGLBERYDPPDRSRRLPRPIINTPMSTLVSEAPLPRLEBPVVEFALLEVER 2406  
2315 TKPICVFMNHSILVSGGWSARGCEVFPNBSHVSQCQNMHTSPVLMDOVSRENGEIL 2374  
2407 TKVCCVFMNHSILVSGGWSARGCELSKRNTHVACQCHTASVAVLNDISRENGEVL 2466  
2375 PLKTLTYVALGVTLLALTLTFFPLTLIRLSNQHGIRNMLTAALGLADVLGLGINAD 2434  
2467 PLKIVTYAVALSTLALTLVFLVSLVRMRSNLHSHKHLAVALFLSOLVFVIGINQRE 2526  
2435 LPPACVIVILHFLVLCFESMALTEALHLYALTEYRDVNNCPMFPYMLGNGVAFIT 2494  
2527 NPLCTVVALHLYITMSTFAMTLVSLHVRMLTEVRNIDTGPMPFYVVGKIPALIV 2586  
2495 GLAVGLDPBGVGNPDCEMLSTVPTLIMSPFAGVPAFAVMSVFLYIILABASCAHQCFE 2554  
2587 GLAVGLDPQGVGNPDCEMLSTODTLIMSPFAGVIGAVIINTVSVLSAVSCQKHHYVG 2646  
2555 KKGPPVGLQSPVAVLLLSATWLLLSVNSDTLLFHYLPATCNCTIQGPPIFLSYVLSK 2614  
2647 KKGIVSLRTAFLLLSATWLLLSATWLLGLAVNRDALSFHYLPFAIPSGVGLFFHCVLQ 2706  
2615 EYRKALK-LACSRKRPDPALTTKST-LTSSVNCSPRYADG-RLYOPVDSAGSLHNS 2670  
2707 EYRKHLKGVGGRKHLHEDATIRATLLTRSLMCTTFGDPGPMLTDLGEBSTASLDSTV 2766  
2671 RS-----GKSQPSYIPLLAREESALNPQGGPGLGDPGSLFLGQDQ 2713  
2767 RDEGIQKGVSSGLVNGSHGEPDASLMPSCXO-----PRG----- 2802  
2714 HBDPTSDSDSLLEDQSGSYASTHSSDSSEEBEEREAAPREGQMSLCPGARLP 2773  
2803 --HDSDSDELST-DEOSSVASHSSDSSEDDGVGABEK-----WDPARG-----A 2845  
2774 LHSTPK-DGPGPGKAPMPDPTGTAKSSGNGAP-----EERLRNGDAL 2818  
2846 VHSITPGDAVANVPAGMPPOSILAESDSEDPGSKPLKTKVSVLHREBQSHRGYR 2905  
2819 SREGSLPLPGSSAQF---HKGILKKKCL---PTISEKS--SLRLPLBQCTGSSRGSS 2869  
2906 PDQESGAARLASQPEPQKGLKKKVTYPPPLTLTBOQLKGLBKCLADGQSPITSR 2965  
2870 ASRGSGRGP---PPRPPRQSIQEOQLNGVMPFAMSIKAGTVDESSGSE 2915  
2966 TSSLSGSGPDCAITVXSPGREGRDHLNGV---AMNVRTGSAQADGSDSE 3012

RESULT 15  
US-10-174-677-77  
Sequence 77, Application US/10174677  
Publication No. US20030190704A1  
GENERAL INFORMATION:  
APPLICANT: Xie, Ting  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR  
FILE REFERENCE: 40716 (IP-012)  
CURRENT APPLICATION NUMBER: US/10/174,677  
CURRENT FILING DATE: 2002-06-19  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 77  
LENGTH: 3014  
TYPE: PR1  
ORGANISM: Homo sapiens  
US-10-174-677-77

Query Match 56.3%; Score 8754; DB 12; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;

Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  
12 TPEPLULLLILLLEPPLLGDPVCRSISNGRSGSCAPMGLCPSSASNLMLYTSR 71  
57 TTRARELLI-----DVGDRGLAGR-KVVSAGRPL-----PLQVLY----- 93  
72 CDDAGTEILGHLVPHRDGLRWCPSEBAHILPPAREGCPMSCRLLIGIHLSPQKLT 131  
94 AASAPTLASRLI-----RAPHLP-----CGAARILCGTGARLC--GALCF 133  
132 P-----EEHCLKA-----PLRQCS-----CKLAQAPG- 155  
134 PVPGCCAAAGSHALAPTLTPACRCPPRPACPPRPICLPBGSSVRLRLCALRRAGA 193  
156 LRAG-----ERSPEESLGRKRRUNVTAP-----QOPSPSYQAT 189  
194 VAVGLALEATAGTSPSPSPPLPPLPBARAGPARARAGTSGRSIKRPMNYQVA 253  
190 VERNOPAGTPVASILAPIDDEGEAGRLHYTMDALFDSRSNOFSLDPVTGAVTTABEIDR 249  
254 LFEENPAGTLILQLAHAYTIEGEBERVSVMGELFDEBSRGYFRIDSATGAVSTDSVDR 313  
250 ETKSTHVRVYTAQDGMRRSALATLTLLTMDNDHDVPEQOEKESLREVLKGYEVL 309  
314 EYKEHVLRVKAVDSTPRSATYITVAVKDTNDSVPFQSEYREKREVRNMLEVEYVL 373  
310 TYRATDGPAPNANILYRLLESGSGSPSEVEIIDPRSGVIRTRGVDRDEEVSYQLTVEA 369  
374 TIRASDRSPINANIRYVLGGAM---DYQLNSSGCVSTRVATLDEBAEAYLLVEA 429  
370 SDQGEDPGRSTTAAVPLSVEDDNDNAPQPSHKRYVVOVEDVTPGAVLYRTASDRDK 429  
430 NQGNRPPLSLATVAVYIEVEDENDNYPQPSQNTVVOVPEBVGLMTAVLRQADBDROG 489  
430 SNAVHYIIMSGNAGOPYLDAQGLDVSPDVEYETKTEKTLRRADGGRRPLSNVSG 489  
490 QNAAIHSTLSGNVAGQYTLHSLGSLIDVINEPDEVDQKSLSKAQDGGRRPLINSBG 549  
490 LVTVOVLINDNAPLFTVSTPQATVLESVPLGYLVLAHQALDADAGDNARLELYLA--- 545  
550 VVSQVLDVNDNEPFLFVSSPQATVLENVPLGYVPHIQAVDADGEMARLHRYLVDTAS 609  
546 ---GVGH-----DPEFTINGTGMISVAALDDEEVDYFSGYEAADHGTALTA 592  
610 TELGGSGAGPKNPATPFPQIHNSSGMITVGCALDEEBEHRHSFGVGAADVHSGPPMSS 669  
593 SASVSVTLVDVNDNPTPOPEYTVRLNEDAAVGTSVYTSVAVDASVITYQITSGNT 652  
670 STSVSITVLDVNDNDPVFTOPTYELRLNEDAAVSSVLTILQARDNANSVITYQITGNT 729  
653 RNRFSITSQSGGGLVALPLDYKLERQYLVATASDGTRODTAQIYVNVTDANTHPVF 712  
730 RNRFPALSSORGGGLITLALPLDYKREQYLVATASDGRSHTAHVLINVTDANTHRPVF 789  
713 QSSHTYVANNEPRPAGTTVVLISATDEDTGERAARTTYMEDSIPFRIDADVGAATTQAE 772  
790 QSSHTYVSEDRPAGTISATLSANDEDTGERAARTTYIQBPVPPFRIDPDSGTWYTWME 849  
773 LDYEQVSVTLATARDNGIIPQKSDPTYLAILVNDVNDNAPQFLDSYQSGSYEDVPPT 832  
850 LDYENQVAVTLTIAQNDGIPQKSDTTLEIILIDANNAQFLMDYFQSGIFEDAPST 909  
833 SVLQISATDRDSGLNGRVYFTTQGGDDGDPFVSTSGIVRTLRRLDRENVAYQVLYAY 892  
910 SILQVSATDRDSGPRGLLYTFQGGDDGDPFYLEPTSGIVRTORRLDRENVAVNLVAL 969  
893 AYDKGMP-PARTPEVTVTVLVDVNDNPPVPRQDEDFVBEENSPLGLAVARTATDPBG 951  
970 AYDKGSPPLPLASAVELQVTLIDINDNAPMFEDDELTFVEZENPFGSVAAKIRADDPBG 1029  
952 TNAQIMQIVENGIPEVFOLDIFSGELTALVDLYEDRPEVYLVQATSAPLVSRATYH 1011  
1030 PNAQIMQIVENGMHFPQDLINDGLDRAVVELDPEVREBYLVVQATSAPLVSRATYHI 1089

QY	1012	RLLDNRNDPVLGNFELLFNNTYVNRSSFPFGAIGRVAPDPIISDLTYSFENGNELS	1071
Db	1090	LLVQNDNBPVLPDPQILFNNTYVNRSSFPFGVIGCIPADHPDVSILNTFYQGNELR	1149
QY	1072	LVLNASTGELKLSALDNNRPLEAMIVLVSDDGHSYTAQCAAPVLTITBMLTHSTTL	1133
Db	1150	LILNDPAIGELQLSLDLNNRPLEAMERSVSDGHSYTAFTLKTVLTITDMLTNSITV	1209
QY	1132	RLSDMSPERFLSPLLGLFIOAAVATLATPPDHVVFNVQRTDAGGHIINLVSVQCP	1191
Db	1210	RLNNMSQKFLSPLLALFVEGVAAVLSTKDVVFENVQNTDV--SSNIINVTPEALLPG	1266
QY	1192	GRGGPRPLPBESDLOERLYNRSLLTIAISQVLPFDNITCLREPCENYMRCSVYLAPDS	1251
Db	1269	GVRG--QFPESBDIQEQIYLNKTLTITSTORVLPFDNITCLREPCENYMRCSVYLAPDS	1326
QY	1252	SAPFIASSSVFRRPIHPVGLRCRCPPTGYCTEVDLCYSRCGPHRCRSBEGYT	1311
Db	1327	SAPFISSITVLFRRPIHPINGLRCKRCPPTGDTCTEITDLCTSDPCANGKCRSHGGYT	1366
QY	1312	CLCRDGYTGEHCERSASRGCTPGVCXKNGCTCVNLVGGPKDCPSGDFFKPYCQVTRS	1371
Db	1387	CECEDPFGEHCERVDARSGRCANCGCKNGCTCVNLIGAFHCVCPCGGEYERBPCEVTRS	1446
QY	1372	FPAPSFIFRGRORFHTLTLSPATKRDGLLYNNGPNKHDPVLALEVIOEQVOTFRS	1431
Db	1447	FPPOSFVIFRGLRQRFHTLTLSPATKRDGLLYNNGPNKHDPVLALEVIOEQVOTFRS	1506
QY	1432	AGESTTVSPVPFGVSDGOWHTVQALKYNRPLEQDTGLPOGSPBOKVAVTVDGCDTGV	1491
Db	1507	AGETTTVAAPKPSGVSDBGRMHSVQVYXNKPNIQHLLPHQPSGEKMAVTVDDCDTGM	1566
QY	1492	ALREGSVLGNVSCAAQCTGGSKSLDITGPIILGGVPLDRESPVYMRKQVYVGCNRIQV	1551
Db	1567	AVRGRKQIGNVSCAAQCTGTSKSLDITGPIILGGVFNLEBEDFPVNRQVPGVCNRIQV	1626
QY	1552	DSRIIDMADFIANNGTVAGCPAKKXVCSNITCHNGTGVNOMDAPSCBCLPAGFGSKCAQ	1611
Db	1627	DGKXVDMAGFIANNGTREGCAARNPFCGRGQNGCTGVNMMNYLCECPILRFGKNCBQ	1666
QY	1612	EMANRPHLGGSLVAMHGLSLPISQPTVLSLMTFRQADGVTLQAITRGSITTLQLREB	1671
Db	1687	AMPHRPLESGSVSWSDNLITISVPTVGLMTFRKEDSVLMEKTSGGSPISFLQITLNN	1746
QY	1672	HVMLSVEGTGLQASLSLRLEPGRANDGMHQAOLAL--GASGCGPCHAL-LSFDPGQOGRAB	1727
Db	1747	YLQREVSHPGSDVESVMLSGLRYVDGBHHLLIELAKVKEBSEKHLYMTLTLDTGMQNK	1806
QY	1728	GNLQPRLHGLHLSNITVGGIDPAGAGVARGFGCLQGVAVSDTBEGVNSLDPHSGESINV	1787
Db	1807	ADIGGMLPGLTVRSVVVGAGSEBDKSVARGFRGCMQGRMGFTPTNVALTMNNALAKRV	1866
QY	1788	EGGSLRPPCCSNPCPANSYCSNMDVSGSCDPEYVDNCTVNCDLNPGCHOSVCTRKP	1847
Db	1867	KDCGCDVDDPCTSSPCPFPNSRGHDAWEDYCYCDKGYTLINCYDAHNLPPCNMGACVRSF	1926
QY	1848	SAPHGYTCECPNYLGPYCTRITDQPCBPGWGHPTQPCNCVDVSKGFPDPCNTKTSGECH	1907
Db	1927	GPSQGYVCECGPSHGYPCENKTLDPCCPRGWMGNVGPCGHCAVSKGFPDPCNTKNGCQ	1966
QY	1908	CKENHYRPRGSPCTCLCDYCPYTGSLSRVCDDEQCCPKREVITGRQDCRCNPPAPVYTTN	1967
Db	1987	CKENYTKLADDTCLPCDCPFGHGSISRCMDATGCAKCPVITGQCCRCNCPAPVAVTTL	2046
QY	1968	GCEVNVYDCPRAIEAGIWPRTTRFGLPAAACCPKGSFCTAVRHCDENHGMLEPNIUNCTS	2027
Db	2047	GCEVYVNGCPAPAFEGIWWPQTKGQAPAAVPCPKGSVGNARHSGSGEKMPLPELFNCTT	2106
QY	2028	ITTBELKGFARLQRNBSGLDSGSGOQALLNRATQHTAGYRGSVDYKVAVQATRLIAH	2087
Db	2107	ISFDVLKPMNKLSRNETQVYDGAARLQVLRALRSATQHTGTLFGADVATAVQLLGHVLOH	2166

QY	2088	ESTORGEGLSTYODVHTEMLRFGSALBLPTANRHEMLQOTEGGJAMLLQHTAEVAASA	214.7
Db	2167	ESWQCGFDLAIATQADPHEBDIVHSGSALLAPATRAAMEQIGRSBGTAQILRLRLEGFSN	222.6
QY	2148	LAQWERTIYLSPEFTIYTPNIVISVYRLDKGNFAGAKIPRIEALGGEOPDLETTYILPES	220.7
Db	2227	VARVNRRTYLRPFYIVANNMLAVDIDPKNFPGARPRPRTTHEEEPRERLESSVSPAD	228.6
QY	2208	VFR----ETPPVVRPAG-----DGAQPEPELARRQRPHELSQGEAVAVITYRT	225.4
Db	2287	FFRPEEKEGBLARPAGRRITTPQTRGPGCTERAPAPISRRGRHDDDAQFVALVITYRT	234.6
QY	2255	LAGLLPHNYOPDKRSLEVPKRPITNTPVSI SVHDBELLPRALDKQVTVQFLLTEER	231.4
Db	2347	LGQLLPERYDPDRSLILPFRPIINTPMVSTLYVSEGAPLPRPLERPVILVEFALLEVER	240.6
QY	2315	TKPICVWMNHSILVSGGWSARGCEVFPNESHVSQOCHMHSFAYIMVSRRENEIL	237.4
Db	2407	TKPVCVWMNHSILVAGGWSARGCELLSNRTHHVAQCSHTTASFAYIMDISRENEGVL	246.6
QY	2375	PLKTLTVVALGVTLAALLTFPEFLTLLRIIRSNQIGRNLITALGLAOLVFLGINOAD	243.4
Db	2467	PLKIVTVAAVVSLSAALLVAFVLSIVRMLRSMNHSIHKLAVALLPLSQLVFVIGINOTE	252.6
QY	2435	LPFACVYIALLHFLYICTFSWALLBLHL YRALTEYRDVNTGPMRYTLMGKVPFIT	249.4
Db	2527	NPFICVYIALLHXYIYVSTPAMTLVBSLHAYRMLTEVRNIDTGPWREYVYVWGMPAIVT	258.6
QY	2495	GLAVGLDPBGXGMPDFCMLSIYDTLMSFAPGPAFAVMSVFTYIILABASCAORQGE	255.4
Db	2587	GLAVGLDPBGXGMPDFCMLSDOTLIMSFPGIGAVIINTVTSVLSAKVSCQRKHYYG	264.6
QY	2555	KKGVSGLQPSFAVLLLSATWMLLALSVNSDTLLFYLTPATCNCIOGPEIFLYSVYLSK	261.4
Db	2647	KKGIVSLTRAFYLLILLISATWMLLGLLAVNRDLSFHYLLPAIFSGLGPFVLLPHCYVLMQ	270.6
QY	2615	EYRKALK-LACSKRSPDPAALTXTST-LTSSVNCPSRYADG--RLYOPYDSDAGSLHSTS	267.0
Db	2707	EYRKHLKGVLGGRKTLHEDSATTRATILLTSLNCNTTGGGPDMLRTLDLBSTASLSOSIV	276.6
QY	2671	RS-----GKSOPSYVLPFLRBSALNPGCGPFGADPGSLFLEGDOQ	271.3
Db	2767	RDEGIQCLGVSGVLNGSHGEPDASLMPRCKD-----PPG-----	280.2
QY	2714	HDPDTSDLSLSIEDDOSGSYASTHSDSZEZEZEZEZEBAALFPEBGQWDSLIFGARLP	277.3
Db	2803	--HDSQSDSELST--DEQSSSYASHSDSDSDDDGVAREK-----WDARG-----A	284.5
QY	2774	LHSTPK--DGGPGRGKAPWPGDFGTALAESGNGAP-----BERLRBNDAL	281.8
Db	2846	VHSTPKDANVANNVPACMPDQSLAESBSEBPSGKPRKVTYKYSVELHREBQSHRGEP	290.5
QY	2819	SREBSGLPLGSSAOP---HKGILKKKCL---PTISEKS--SILRLPLBQCTGSSRGSS	286.9
Db	2906	PDQESGGAATLASSQPEBGKGLIKKVKVYPPPLITLTBQLKGLRKLADCEQSPTSR	296.5
QY	2870	ASBESRGCP-----PRRPPRQSIQOBQUNGMPILAMSIKGTVEDSDSGSE	291.5
Db	2966	TBSLGSGBPCPACITVNSPGREPRDHLNGV--ANNVTSASAQADGSDSE	301.2

Search completed: February 11, 2004, 16:05:02  
Job time : 79 secs

Db 2107 ISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQHTGTLEGNDRVATAYQLLGHVLQH 2166

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 11, 2004, 15:50:43 ; Search time 61 Seconds  
(without alignments)  
7605.865 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 15545  
Sequence: 1 MRSPTGVPLPTPPPLLL.....AGTVDESSGSEFLFNFIAH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 156726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq.130un03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15545	100.0	2923	22	AAU07054 Human Flamingo pro
2	15545	100.0	2923	23	AAU07054 Human Flamingo pro
3	15545	100.0	2923	24	AAU07054 Human Flamingo pro
4	15518.5	99.8	2956	22	AAU07053 Human GPCR CELSR2
5	15393	99.0	2894	24	AAU07053 Human GPCR CELSR2
6	15279	98.3	2936	23	AAU074826 Human MDP1 polypep
7	13399	86.2	2560	22	ABU1404 Human REPT 9 prot
8	12541.5	80.7	2405	21	AAU07192 Human FLAMINGO 1 h
9	8754	56.3	3014	22	AAU06533 Human novel cytoke

10	8754	56.3	3014	22	AAU02196	Seven-pass transme
11	8754	56.3	3014	24	ABP81979	Human GPCR CELSR1/
12	8709	56.0	3028	22	AAU08586	Human NOV7 protein
13	8495	54.6	2707	18	AAU07161	Mouse receptor ME2
14	7816.5	50.3	4115	23	ABP33570	Human NOV2 protein
15	7763	49.9	3298	22	ABU03657	Human extracellular
16	4817.5	31.0	3606	22	ABU02595	Drosophila melanog
17	4369.5	28.1	1572	18	AAU07160	Mouse receptor ME2
18	4093	26.3	1460	24	ABP81741	Human CELSR3 prote
19	3641	23.4	717	21	AAU06721	Human prostate can
20	2958	19.0	568	19	AAU01244	Human calcitonin r
21	2937	18.9	568	23	ABU09665	Human polypeptide
22	2900.5	18.7	565	24	ABU07074	Human breast cance
23	2425.5	15.6	1554	22	ABU11890	Human protocadheri
24	1814.5	11.7	470	24	ABP75877	Human secretory po
25	1712.5	11.0	4903	24	ABU12072	Human NOV17a CG928
26	1534.5	9.9	5147	22	ABU09831	Drosophila melanog
27	1516.5	9.8	4590	23	ABU06808	Human pancreatic c
28	1516.5	9.8	4590	24	AAU06790	Human cadherin (CA
29	1516.5	9.8	4590	24	AAU06791	Human cadherin (CA
30	1516.5	9.8	4591	22	ABU02977	Novel human diagno
31	1477.5	9.5	4643	22	ABU11609	Drosophila melanog
32	1419.5	9.1	3097	22	ABU02967	Drosophila melanog
33	1413	9.1	4555	23	AAU02106	Rat fat 3 protein
34	1336.5	8.6	3909	23	ABP43661	mRNA for hPat prot
35	1320.5	8.5	4349	23	AAU09940	Human protocadheri
36	1320.5	8.5	4349	23	ABU07540	Novel human protei
37	1320.5	8.5	4349	24	AAU06792	Human cadherin (CA
38	1262.5	8.1	3298	23	ABU05430	Human dachshous pro
39	1262.5	8.1	3298	24	AAU06793	Human cadherin (CA
40	1260.5	8.1	4263	23	ABU07541	Novel human protei
41	1217.5	7.8	3354	24	AAU06788	Human cadherin (CA
42	1217.5	7.8	3354	24	AAU06789	Human cadherin (CA
43	1210.5	7.8	3217	23	AAU04825	Human REPT 8 prot
44	1207	7.8	3503	22	ABU06499	Drosophila melanog
45	1137.5	7.3	1783	22	ABU03321	Drosophila melanog

#### ALIGNMENTS

RESULT 1	AAU07054	standard; Protein; 2923 AA.
ID	AAU07054	standard; Protein; 2923 AA.
AC	AAU07054	
XX		
DT	24-OCT-2001	(first entry)
XX		
DE	Human Flamingo protein encoded by cDNA splice variant.	
XX		
KW	Flamingo; human; splice variant; G-protein coupled receptor; diabetes;	
KW	signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;	
KW	anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;	
KW	obesity; hypertension; hypertension; urinary retention; angina pectoris;	
KW	myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;	
KW	benign prostatic hyperplasia; psychotic disorder; neurological disorder;	
KW	manic depression; delirium; dementia; severe mental retardation;	
KW	Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;	
KW	antifungal; antiviral; antiprotocadherin; anti-HIV; anorectic; antianorectic;	
KW	antiparkinsonian; cardant; cerebroprotective; neuroprotective;	
KW	antidepressant; anticonvulsant; antisense therapy; gene therapy.	
OS	Homo sapiens.	
XX		
PM	WO200161003-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	19-FEB-2001; 2001WO-GB00680.	
XX		
PR	19-FEB-2000; 2000GB-0004196.	
XX		

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX Testa TT;  
 XX WPI; 2001-502792/55.  
 DR N-PSDB; AAS11678.  
 XX  
 PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 PS  
 PS Claim 2; Page 33-34; 66pp; English.  
 XX  
 XX The sequence represents a human Flamingo polypeptide encoded by a cDNA  
 CC splice variant polynucleotide. Flamingo is a member of the G-protein  
 CC coupled receptor family, which is involved in signal transduction  
 CC pathways. By screening to identify compounds that stimulate or inhibit  
 CC the function or level of the protein, treatments can be developed for  
 CC various diseases and bacterial, fungal, protozoan and viral infections,  
 CC including HIV, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, angina pectoris, myocardial infarction, stroke,  
 CC ulcers, allergies and benign prostatic hypertrophy. Also treatable are  
 CC psychotic and neurological disorders such as anxiety, schizophrenia,  
 CC manic depression, delirium, dementia, severe mental retardation,  
 CC Huntington's disease and Gilles de la Tourette's syndrome.  
 CC  
 XX  
 XX Sequence 2923 AA;  
 SQ  
 Query Match 100.0%; Score 15545; DB 22; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 LDVNDNNPTFQPEYTVLNEADAAGVTSVTVSAVDRDASHVITVQITSGNTRNRFSTIS 660  
 DB 601 LDVNDNNPTFQPEYTVLNEADAAGVTSVTVSAVDRDASHVITVQITSGNTRNRFSTIS 660  
 QY 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVTTDANTRPVFOSSHVTVN 720  
 DB 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVTTDANTRPVFOSSHVTVN 720  
 QY 721 VVEDRPAGTVVLIATDEDEGENARITYFMEDSIPQRIADTGAVTQOAEIYEDDOVS 780  
 DB 721 VVEDRPAGTVVLIATDEDEGENARITYFMEDSIPQRIADTGAVTQOAEIYEDDOVS 780  
 QY 781 YTLATTAADNGIPQKSDTYLLEILVNDVNDNAPOELRDSYOGSYVEDVPFTSVLQISAT 840  
 DB 781 YTLATTAADNGIPQKSDTYLLEILVNDVNDNAPOELRDSYOGSYVEDVPFTSVLQISAT 840  
 QY 841 DSDSLNGRVFTTFCGGDDGGDPFVVESTSGVTRLRDLBENNAQYTLRAYADKGMPP 900  
 DB 841 DSDSLNGRVFTTFCGGDDGGDPFVVESTSGVTRLRDLBENNAQYTLRAYADKGMPP 900  
 QY 901 ARTPMETVTVLDVNDNPPVFEODEPQVFEENSPIGLAVARVLTATDPDEGTNAQIMYQI 960  
 DB 901 ARTPMETVTVLDVNDNPPVFEODEPQVFEENSPIGLAVARVLTATDPDEGTNAQIMYQI 960  
 QY 961 VEGNIPEVFQDLDFSGELTALVDLDYDRPREVVIQATSAELVSRAVTVHVLDRNDNP 1020  
 DB 961 VEGNIPEVFQDLDFSGELTALVDLDYDRPREVVIQATSAELVSRAVTVHVLDRNDNP 1020  
 QY 1021 PVLGNFELLFNNTYNNRSSSPFGAIGVPAHDPIISLITYSPFRGNELSVILNASTG 1080  
 DB 1021 PVLGNFELLFNNTYNNRSSSPFGAIGVPAHDPIISLITYSPFRGNELSVILNASTG 1080  
 QY 1081 ELKLSRALDNNRPLAISMVLSVDSGHSVTAQALRVITITBEMLTHSITTLRLJEDMSER 1140  
 DB 1081 ELKLSRALDNNRPLAISMVLSVDSGHSVTAQALRVITITBEMLTHSITTLRLJEDMSER 1140  
 QY 1141 FLSPILGLFTQVAATLTATPPDHVVVFNQRTDAPGSHILNLSVQPPGCGPPFL 1200  
 DB 1141 FLSPILGLFTQVAATLTATPPDHVVVFNQRTDAPGSHILNLSVQPPGCGPPFL 1200  
 QY 1201 PSEDLOERLYNRSLLTIAISORVLPDPDNTCLRPCCNMYRCVTLAPDSGAPFIASS 1260  
 DB 1201 PSEDLOERLYNRSLLTIAISORVLPDPDNTCLRPCCNMYRCVTLAPDSGAPFIASS 1260  
 QY 1261 VLFREIHPVGLRCRCPFGTGDYCEFEVDLCYSRPGCPHRCRREGGYTCLCRDGYTG 1320  
 DB 1261 VLFREIHPVGLRCRCPFGTGDYCEFEVDLCYSRPGCPHRCRREGGYTCLCRDGYTG 1320  
 QY 1321 EHCYVSANSGRCTPGVCNKGCTCVNLVYGFKCDPCSGDPFKRYCQVTRSPAFSFTTF 1380  
 DB 1321 EHCYVSANSGRCTPGVCNKGCTCVNLVYGFKCDPCSGDPFKRYCQVTRSPAFSFTTF 1380  
 QY 1381 RGLRORFHTLALSAATBERDGLLYNGRFNKKHDFVLEVLQEOVLTFSGESTTVYS 1440  
 DB 1381 RGLRORFHTLALSAATBERDGLLYNGRFNKKHDFVLEVLQEOVLTFSGESTTVYS 1440  
 QY 1441 PVPVGSVSDGQWHTYQIKTYNRPILGOTGLPOGSEBOKYAVTVYDSCDTGVALRGSYLG 1500  
 DB 1441 PVPVGSVSDGQWHTYQIKTYNRPILGOTGLPOGSEBOKYAVTVYDSCDTGVALRGSYLG 1500  
 QY 1501 NYSCAAQGTQGGSKSLDTGELLGCVPDLPESPPVMRQOVGCMRLQVDSRIIDMAD 1560  
 DB 1501 NYSCAAQGTQGGSKSLDTGELLGCVPDLPESPPVMRQOVGCMRLQVDSRIIDMAD 1560  
 QY 1561 FIANNGTVPGPCAKKNVCDNSNCHNGGTCVNOMDAPSCBCLGFGGSKCAQEMANPQHTL 1620  
 DB 1561 FIANNGTVPGPCAKKNVCDNSNCHNGGTCVNOMDAPSCBCLGFGGSKCAQEMANPQHTL 1620  
 QY 1621 GSSLVAMHGLSLPISQPMYLSIMFRTQADGVLLQAITRGSTITLQLRBGMVLSVBGT 1680  
 DB 1621 GSSLVAMHGLSLPISQPMYLSIMFRTQADGVLLQAITRGSTITLQLRBGMVLSVBGT 1680  
 QY 1681 GLQASLRLEPGRANDGWMHNAQLALGASGGFGHAILSTDYQOQRAEGNLGRLGHLHS 1740



```

Db 1681 GLOASSLBERADGWHHQAALGASGGGHAHLISFDYQOQRAEENLGRHLGHLIS 1740
Qy 1741 NITWGGIPGAGVARGFRGCIQGVASDTPGVNSLDPESHESINVEQCSLPDPCSN 1800
Db 1741 NITWGGIPGAGVARGFRGCIQGVASDTPGVNSLDPESHESINVEQCSLPDPCSN 1800
Qy 1801 PCPANSYCSNDWDSYSCSDPGYTDNCTNCDLNPCHQSVCTKPSAPHYGTCECPN 1860
Db 1801 PCPANSYCSNDWDSYSCSDPGYTDNCTNCDLNPCHQSVCTKPSAPHYGTCECPN 1860
Qy 1861 YLGPYCEFRIDOPCRGMMGHPTGSPCNDVSKGPDPCNTSGSCHKEHHYRPPGSP 1920
Db 1861 YLGPYCEFRIDOPCRGMMGHPTGSPCNDVSKGPDPCNTSGSCHKEHHYRPPGSP 1920
Qy 1921 CLUCDCYPTGSLSRVCDPEDGQPCPKPGVIGRCDCRCNPFALVTTCNCEVYDSCPRAI 1980
Db 1921 CLUCDCYPTGSLSRVCDPEDGQPCPKPGVIGRCDCRCNPFALVTTCNCEVYDSCPRAI 1980
Qy 1981 EAGIWPRTTRFGLPAAPCPKGSFGTAVRHCDHRGMLPMLFNCTSI7FSELKGPAREL 2040
Db 1981 EAGIWPRTTRFGLPAAPCPKGSFGTAVRHCDHRGMLPMLFNCTSI7FSELKGPAREL 2040
Qy 2041 QRNESGLDSGRSQOALILRNATQHTAGYFSDVAVAYQATRLAHBSTORGFGLSATQ 2100
Db 2041 QRNESGLDSGRSQOALILRNATQHTAGYFSDVAVAYQATRLAHBSTORGFGLSATQ 2100
Qy 2101 DVHFTENILRVGSLALDTPANKRMELIOQTEGTAMLQHYEAYASALANMRHTYLSPF 2160
Db 2101 DVHFTENILRVGSLALDTPANKRMELIOQTEGTAMLQHYEAYASALANMRHTYLSPF 2160
Qy 2161 TIVTPNIVISVVRLDKGNPAGAKLPRYBALRGEOPDLETTVIIPEVFRETPPVVRDAG 2220
Db 2161 TIVTPNIVISVVRLDKGNPAGAKLPRYBALRGEOPDLETTVIIPEVFRETPPVVRDAG 2220
Qy 2221 PGEAOPEBELARBORRHPELSQGEAVASVIIYRITAGLIPNNYOPDKSLAVPRPIINT 2280
Db 2221 PGEAOPEBELARBORRHPELSQGEAVASVIIYRITAGLIPNNYOPDKSLAVPRPIINT 2280
Qy 2281 FVVISIVHDEBELLPALDKPVTVQFRLLETERTKPICVFMNHSILVSGTGSARCE 2340
Db 2281 FVVISIVHDEBELLPALDKPVTVQFRLLETERTKPICVFMNHSILVSGTGSARCE 2340
Qy 2341 VVFRNESVSCQCNMTSPAVLMDVSRRENGEILPLKLTIVVALGVTLAALLTFPFLT 2400
Db 2341 VVFRNESVSCQCNMTSPAVLMDVSRRENGEILPLKLTIVVALGVTLAALLTFPFLT 2400
Qy 2401 LRIIRSNQHGIRBNLTALGLAQLVFLGINQADLPFACTVIALILHFLYCTFSMALLB 2460
Db 2401 LRIIRSNQHGIRBNLTALGLAQLVFLGINQADLPFACTVIALILHFLYCTFSMALLB 2460
Qy 2461 ALHLYRALTEVRDVTGPMRFYMLGMGVPAFITGLAVGLDPBGYNDFCMLSYDTLI 2520
Db 2461 ALHLYRALTEVRDVTGPMRFYMLGMGVPAFITGLAVGLDPBGYNDFCMLSYDTLI 2520
Qy 2521 WSPFAGPVAFAVMSVFLYTLAARASCAAQROGFEXKGVSGIQSFVALLLSATMTLAL 2580
Db 2521 WSPFAGPVAFAVMSVFLYTLAARASCAAQROGFEXKGVSGIQSFVALLLSATMTLAL 2580
Qy 2581 LSVNSDTLFHYLPTATCNCIGPFIPLSYVVLSTKEVRKALCLAGRKSPDPALTSTL 2640
Db 2581 LSVNSDTLFHYLPTATCNCIGPFIPLSYVVLSTKEVRKALCLAGRKSPDPALTSTL 2640
Qy 2641 TTSYNCPSPYADGRLYQPYGDSAGLSHSTRSGSQSPYSIPFLLRBSALNPGQPPGLG 2700
Db 2641 TTSYNCPSPYADGRLYQPYGDSAGLSHSTRSGSQSPYSIPFLLRBSALNPGQPPGLG 2700
Qy 2701 DPGSLFLEGQOQOHPDPTSDPSLSLEDQSGSIASTHSSPSEEBEEREEAAPPQOG 2760
Db 2701 DPGSLFLEGQOQOHPDPTSDPSLSLEDQSGSIASTHSSPSEEBEEREEAAPPQOG 2760
Qy 2761 WDSLILGPAERLPLHSTPKDGPBGKAPWPGDFTTAKESSGNGAPERRLRNGDALSR 2820

```

```

Db 2761 WDSLILGPAERLPLHSTPKDGPBGKAPWPGDFTTAKESSGNGAPERRLRNGDALSR 2820
Qy 2821 EGSIGPLPGSSAOPHKGLKKKCLPTISEKSLRLPLBOCTGSSRSGSASGSGGPP 2880
Db 2821 EGSIGPLPGSSAOPHKGLKKKCLPTISEKSLRLPLBOCTGSSRSGSASGSGGPP 2880
Qy 2881 RPPRQSLQEQNLGMPFIAMSIRKAGTVDEDSGSEFLFPFNLIH 2923
Db 2881 RPPRQSLQEQNLGMPFIAMSIRKAGTVDEDSGSEFLFPFNLIH 2923

RESULT 2
AAM50866
ID AAM50866 standard; Protein; 2923 AA.
XX
AC AAM50866;
XX
DT 07-MAY-2002 (first entry)
XX
DE Cadherin EGF IAG seven-pass G-type receptor 2, basal cell marker.
XX
KW Cadherin EGF IAG seven-pass G-type receptor 2; receptor;
KW EGF-like domain, multiple 2; basal cell; marker; breast cancer;
KW tumour; diagnosis; prognosis; classification; human; gene therapy.
XX
OS Homo sapiens.
XX
PN MOJ00208765-A2.
XX
PD 31-JAN-2002.
XX
PE 26-JUL-2001; 2001MO-US23843.
XX
PR 26-JUL-2000; 2000US-220967P.
XX
PA (STRD ) UNIV STANFORD.
XX
FA (GENO-) APPLIED GENOMICS INC.
XX
PI Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
PI Seltz R;
XX
DR WPI; 2002-206094/26.
XX
PT Classifying and creating breast tumours associated with the expression
PT and activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14
PT domain, Multiple 2 -
XX
PS Claim 3, Fig 1C; 741pp; English.
XX
CC The present sequence is that of human cadherin EGF IAG seven-pass
CC G-type receptor 2, also known as EGF-like domain, multiple 2.
CC The invention provides new reagents and methods for the management
CC (e.g. detection, classification, provision of diagnostic and
CC prognostic information, treatment, etc.) of breast cancer. CDNA
CC microarray technology was used to identify genes whose expression
CC profile across a large group of tumour samples correlated with that
CC of cytokeratin 5 and cytokeratin 17, markers for basal cells of the
CC normal mammary lactation gland. Tumours that express cytokeratin
CC 5/6 and/or 17 have a poor prognosis relative to tumours overall.
CC The basal marker genes identified included those encoding cadherin
CC 3 (see AAM50864), matrix metalloproteinase 14 (see AAM50865) and
CC cadherin EGF IAG seven-pass G-type receptor 2. Detection of these
CC markers allows classification of a tumour sample as belonging to a
CC basal tumour subclass. By providing reagents that may reliably be
CC used to classify tumours as belonging to a basal subclass, the
CC invention enables a variety of methods for improving therapeutic
CC options for patients with breast cancer and for individualising
CC therapy. Potential therapeutic agents include compounds that
CC modulate breast basal cell marker genes or that modulate
CC polypeptides encoded by these genes. In particular, therapeutic
CC agents include antibodies that specifically bind to the basal
CC marker proteins and which recognise basal cells of normal mammary

```

CC lactation glands. The antibodies can be used either by themselves  
 CC or when conjugated to, or delivered with, another molecule such as  
 CC a toxic compound. The likelihood that cadherin EGF lig seven-pass  
 CC G-type receptor 2 is membrane-bound makes it an attractive candidate  
 CC for antibody therapeutics.

XX Sequence 2923 AA:

Query Match 100.0%; Score 15545; DB 23; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSFATGVPLEPPPLLLLLLLLLLPPPLGQVGCRLSGRSGSCGAPMGLCPG 60  
 DB 1 MRSFATGVPLEPPPLLLLLLLLLLPPPLGQVGCRLSGRSGSCGAPMGLCPG 60  
 QY 61 SASNMLYNSRCDAGTELTGHLVPHHDGLRWCPSESAHPLPPAPGCGMCSRLGIG 120  
 DB 61 SASNMLYNSRCDAGTELTGHLVPHHDGLRWCPSESAHPLPPAPGCGMCSRLGIG 120  
 QY 121 GHLSPOGKLTLPREHCLKAPRLRCSCKLAQAPGAPGERSPEBSIGGRKXNVTAPO 180  
 DB 121 GHLSPOGKLTLPREHCLKAPRLRCSCKLAQAPGAPGERSPEBSIGGRKXNVTAPO 180  
 QY 181 POPSYQATVPENQAPGTPVASLRALDPGEAGRLBYTMDALFDSRNOFSLDPVTA 240  
 DB 181 POPSYQATVPENQAPGTPVASLRALDPGEAGRLBYTMDALFDSRNOFSLDPVTA 240  
 QY 241 VTTAEELDEETKSTHYFRYTAQDHGMPRSALATLTTLVTDTMDHPVPEQGYKSLAE 300  
 DB 241 VTTAEELDEETKSTHYFRYTAQDHGMPRSALATLTTLVTDTMDHPVPEQGYKSLAE 300  
 QY 301 NLEVGVEVLTARLTDAPPNANILYRLLEGSGSPSEVEIDPRSGVLRTRGPVDRERY 360  
 DB 301 NLEVGVEVLTARLTDAPPNANILYRLLEGSGSPSEVEIDPRSGVLRTRGPVDRERY 360  
 QY 361 ESYQLTVEASDQGRDPGRSTTAAPVLSVEDDNDNAPQESKRYVYVQVEDVTGAPVLR 420  
 DB 361 ESYQLTVEASDQGRDPGRSTTAAPVLSVEDDNDNAPQESKRYVYVQVEDVTGAPVLR 420  
 QY 421 VTSDDKGSNAVYHSIMSGNARGQFYLDQFGALDVSPLDYETTKETTLFVRADQDG 480  
 DB 421 VTSDDKGSNAVYHSIMSGNARGQFYLDQFGALDVSPLDYETTKETTLFVRADQDG 480  
 QY 481 RPLSNVSGIYTVQVLDINDNAPIFVSTPQATVLESPLGYLVLAQALDADAGNARL 540  
 DB 481 RPLSNVSGIYTVQVLDINDNAPIFVSTPQATVLESPLGYLVLAQALDADAGNARL 540  
 QY 541 EYRLAGVGHDPPTINNGTGMISVAELDRREVDFYSPGVABDHGTALTAGASVSTV 600  
 DB 541 EYRLAGVGHDPPTINNGTGMISVAELDRREVDFYSPGVABDHGTALTAGASVSTV 600  
 QY 601 LDVNDNPPFTQPEYVRLNEDAVGTSVTVSAVDAHVSTYQITSGNTNRPSITS 660  
 DB 601 LDVNDNPPFTQPEYVRLNEDAVGTSVTVSAVDAHVSTYQITSGNTNRPSITS 660  
 QY 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVVTADANTHRFVPOSSHYYN 720  
 DB 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVVTADANTHRFVPOSSHYYN 720  
 QY 721 VNEDRPAGITVVLISATDEBTGENARITYEMEDSIPQFRIDADTGAVTQAEILDYEOVS 780  
 DB 721 VNEDRPAGITVVLISATDEBTGENARITYEMEDSIPQFRIDADTGAVTQAEILDYEOVS 780  
 QY 781 YTLAATRRDNGIPQKSTYTLLELVNDVNDNAPQFLDSYQGSYVEVPPFTSVLQISAT 840  
 DB 781 YTLAATRRDNGIPQKSTYTLLELVNDVNDNAPQFLDSYQGSYVEVPPFTSVLQISAT 840  
 QY 841 DRDSGLNGRVYTYFGGDDGDFIVESTSGIVTLERLDRBNVAQYVLAAYVADKMP 900  
 DB 841 DRDSGLNGRVYTYFGGDDGDFIVESTSGIVTLERLDRBNVAQYVLAAYVADKMP 900  
 QY 901 ARTPEMTVTVLVDVNDNPPVFEODEFVFEENSPIGLAVARVTAIDPDGCTNAQIMYQI 960

DB 901 ARTPEMTVTVLVDVNDNPPVFEODEFVFEENSPIGLAVARVTAIDPDGCTNAQIMYQI 960  
 QY 961 VEGNIPEVPOLDIPSGELTALVDLYEDRPREVVIYQATSAPIVSARATVHRLDRNMP 1020  
 DB 961 VEGNIPEVPOLDIPSGELTALVDLYEDRPREVVIYQATSAPIVSARATVHRLDRNMP 1020  
 QY 1021 PVLGNFELLFNNYVYNNRSSPPGALIGRVPAHDPDISLTYSPFRGNELSVLNLASTG 1080  
 DB 1021 PVLGNFELLFNNYVYNNRSSPPGALIGRVPAHDPDISLTYSPFRGNELSVLNLASTG 1080  
 QY 1081 EIKLSRALDNNRPLEAIVSVSDGSHVTAQALRVITIIDEMLTHTSITTLRLBMSPER 1140  
 DB 1081 EIKLSRALDNNRPLEAIVSVSDGSHVTAQALRVITIIDEMLTHTSITTLRLBMSPER 1140  
 QY 1141 FLSPILGLFIQAVATLTTPPDHVVVFNVDTPARAGHILNVSLVSGOPPGPGGAPPL 1200  
 DB 1141 FLSPILGLFIQAVATLTTPPDHVVVFNVDTPARAGHILNVSLVSGOPPGPGGAPPL 1200  
 QY 1201 PSEDLOERLYNRSLLTAISAQRVLPFDNCLAREPCENYRCVSLRFDSSAPFIASSS 1260  
 DB 1201 PSEDLOERLYNRSLLTAISAQRVLPFDNCLAREPCENYRCVSLRFDSSAPFIASSS 1260  
 QY 1261 VLFPRTHVGGILRCRCPGFTGDIYCTEVDLCYSRPPGPHRCNSRSGYTLCRDGYTG 1320  
 DB 1261 VLFPRTHVGGILRCRCPGFTGDIYCTEVDLCYSRPPGPHRCNSRSGYTLCRDGYTG 1320  
 QY 1321 EHCYSASGRCTPCVCKNGGCVNLLVSGFKCDPSPGDFEPCVQVTRSPASFTTF 1380  
 DB 1321 EHCYSASGRCTPCVCKNGGCVNLLVSGFKCDPSPGDFEPCVQVTRSPASFTTF 1380  
 QY 1381 RGLRQRFHTLALSPATEKRDGLLYNGRPNKHDVALLVYQEOVOLTFPSAGBSTTVS 1440  
 DB 1381 RGLRQRFHTLALSPATEKRDGLLYNGRPNKHDVALLVYQEOVOLTFPSAGBSTTVS 1440  
 QY 1441 PVPBGVSDGQWHTYQLYTKNKPILLGOTGLPQGBEOKVAYVTTDGCCTGVALRGSYLG 1500  
 DB 1441 PVPBGVSDGQWHTYQLYTKNKPILLGOTGLPQGBEOKVAYVTTDGCCTGVALRGSYLG 1500  
 QY 1501 NYSCAOQGGGSKSLDTPGLLGGVPLPESFPVMRQPVGCMRLQVDSRIIDMAD 1560  
 DB 1501 NYSCAOQGGGSKSLDTPGLLGGVPLPESFPVMRQPVGCMRLQVDSRIIDMAD 1560  
 QY 1561 FIANNGTVPQCPAKKNVDSNTCHNGGTCVQWMAFSCBCLPFGSGKCAQEMANPOHFL 1620  
 DB 1561 FIANNGTVPQCPAKKNVDSNTCHNGGTCVQWMAFSCBCLPFGSGKCAQEMANPOHFL 1620  
 QY 1621 GSSLVAMHGLSLPIQOPWTLIMFRTROADGVLLQAITRGRSTITLQLEBGMVLSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPIQOPWTLIMFRTROADGVLLQAITRGRSTITLQLEBGMVLSVEGT 1680  
 QY 1681 GLQASSLRLEPRANDGMHNAQLALGASGGGHAITSPDYQQAEGNLGRLGLHLS 1740  
 DB 1681 GLQASSLRLEPRANDGMHNAQLALGASGGGHAITSPDYQQAEGNLGRLGLHLS 1740  
 QY 1741 NITVGGIPQAGVARGRCLOGVRSVDPBGVNSLDPSHGESINVEQCSGLPDPCCSN 1800  
 DB 1741 NITVGGIPQAGVARGRCLOGVRSVDPBGVNSLDPSHGESINVEQCSGLPDPCCSN 1800  
 QY 1801 PCPANSYSGNDMDSYSCSDPQYVGNCTNVDLNPCHQSVCTKRPSPAPHYTCPCPN 1860  
 DB 1801 PCPANSYSGNDMDSYSCSDPQYVGNCTNVDLNPCHQSVCTKRPSPAPHYTCPCPN 1860  
 QY 1861 YLGPYCESTRIDQPCRGWGHPTCPGCPNCVDSKGPDPDCKNTSGCHCKENYRPPGSP 1920  
 DB 1861 YLGPYCESTRIDQPCRGWGHPTCPGCPNCVDSKGPDPDCKNTSGCHCKENYRPPGSP 1920  
 QY 1921 CLLCDCTPTGSLSRVCDPBDGQCPCKPGVYGRQCDRCNDPFAVETTINGCEVYDSCPRAI 1980  
 DB 1921 CLLCDCTPTGSLSRVCDPBDGQCPCKPGVYGRQCDRCNDPFAVETTINGCEVYDSCPRAI 1980  
 QY 1981 BAGIWWPRTFRGLPAAAPCPKSGPGTAARHCOEHGMLPPLNFCNCTSTFSELKGPABEL 2040

```

Db      1981  EAGIWMRTRFGLPAAPCPKSGFTAVHCDHRCMLPNLFCNTSTFSEIKGAEKL 2040
Qy      2041  QNREGSLDGRSQQLALLRNATQTAGYFGSDVAVAYOLAELLAHSTORFGLSATQ 2100
Db      2041  QNREGSLDGRSQQLALLRNATQTAGYFGSDVAVAYOLAELLAHSTORFGLSATQ 2100
Qy      2101  DVHFTNLRVSGALLDTANKHWEILIOCTEGTALLOHYEVYASALANONRHITLSP 2160
Db      2101  DVHFTNLRVSGALLDTANKHWEILIOCTEGTALLOHYEVYASALANONRHITLSP 2160
Qy      2161  TTVTNNIVISVVRDLKGNFAGALPRYBALRGEQPDLETTVILPESVRETPPVVRPA 2220
Db      2161  TTVTNNIVISVVRDLKGNFAGALPRYBALRGEQPDLETTVILPESVRETPPVVRPA 2220
Qy      2221  PGEAOPPEELARORRHPELSOGEAVASVITYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
Db      2221  PGEAOPPEELARORRHPELSOGEAVASVITYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
Qy      2281  PVVSIIVHDEBELPVALDKPVTVQRLLETERTKPICVFNHSLIVSGTGMARCE 2340
Db      2281  PVVSIIVHDEBELPVALDKPVTVQRLLETERTKPICVFNHSLIVSGTGMARCE 2340
Qy      2341  VVFRNESHVSCCNHMTSEPAVIMDVSRRENGELLPLKTLTYVALGVTLAALLTPPELT 2400
Db      2341  VVFRNESHVSCCNHMTSEPAVIMDVSRRENGELLPLKTLTYVALGVTLAALLTPPELT 2400
Qy      2401  LRLNSNQGIRRNLTALALQVFLGGINQADLPACTVIALHFLYLCFTSFALLE 2460
Db      2401  LRLNSNQGIRRNLTALALQVFLGGINQADLPACTVIALHFLYLCFTSFALLE 2460
Qy      2461  ALHLRALTRVDVMTGPMRFYMGVPAITIGLAVLDRENGNPRCMLSTYDTLI 2520
Db      2461  ALHLRALTRVDVMTGPMRFYMGVPAITIGLAVLDRENGNPRCMLSTYDTLI 2520
Qy      2521  WSPFAGVAVASVFLYTLAARASCAQROGFEKKGPVSGQPSFAYVILLSATMTAL 2580
Db      2521  WSPFAGVAVASVFLYTLAARASCAQROGFEKKGPVSGQPSFAYVILLSATMTAL 2580
Qy      2581  LSVNSDTLFFHFLFATCNCIOGPFIFLSTYVLSKEVRKALKACSRKSPDPAITTKSTL 2640
Db      2581  LSVNSDTLFFHFLFATCNCIOGPFIFLSTYVLSKEVRKALKACSRKSPDPAITTKSTL 2640
Qy      2641  TSSVNCPSPYADGRLYOPYGDAGSLHSTRSGKOPSTIPLLEESALNPGQPPGLG 2700
Db      2641  TSSVNCPSPYADGRLYOPYGDAGSLHSTRSGKOPSTIPLLEESALNPGQPPGLG 2700
Qy      2701  DPGSLFLBGOQDQHPDTSDDLLEDDQSGSYASTHSSDBEEREBEEREAAPRGOG 2760
Db      2701  DPGSLFLBGOQDQHPDTSDDLLEDDQSGSYASTHSSDBEEREBEEREAAPRGOG 2760
Qy      2761  WDSLIGPAGAEPLJHSTPKDGGPGGKAPWPGDFTTAKSSGNGAPBEERLENGDALSR 2820
Db      2761  WDSLIGPAGAEPLJHSTPKDGGPGGKAPWPGDFTTAKSSGNGAPBEERLENGDALSR 2820
Qy      2821  BBSLGLPLPGSSAQHPKGLIKKCLPTISKSLALLPLEOCTGSSRGSSASGSGGPP 2880
Db      2821  BBSLGLPLPGSSAQHPKGLIKKCLPTISKSLALLPLEOCTGSSRGSSASGSGGPP 2880
Qy      2881  RPPRQSLQOBOLNGVPIAMSIKAGTVDESSGSEFLPFNFLLH 2923
Db      2881  RPPRQSLQOBOLNGVPIAMSIKAGTVDESSGSEFLPFNFLLH 2923

```

```

KW      G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW      G protein-coupled receptor modulator; antibody; immune-related disease;
KW      growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW      immunological-related cell proliferative disease; autoimmune disease;
KW      Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW      osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW      graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW      psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW      mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW      hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW      ulcer.
XX      Homo sapiens.
XX      WO200261087-A2.
XX      08-AUG-2002.
XX      19-DEC-2001; 2001WO-US50107.
XX      19-DEC-2000; 2000US-257144P.
XX      (LIFE-) LIFESPAN BIOSCIENCES INC.
XX      Burner GC, Roush CL, Brown JP;
XX      WPI, 2003-046718/04.
XX      N-PSDB; AB242868.
XX      New isolated antigenic peptides e.g., for G protein-coupled receptors
XX      (GPCR), useful for diagnosing and designing drugs for treating
XX      conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
XX      cancer or autoimmune diseases
XX      PS
XX      Disclosure; Fig 1; 523pp; English.
XX      The present invention describes antigenic peptides (1) comprising:
XX      (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
XX      acids. Also described: (1) an assay for the detection of a particular
XX      G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
XX      and (2) an isolated antibody having high specificity and high affinity
XX      or avidity for a particular GPCR. (1) can be used as GPCR modulators and
XX      in gene therapy. The antigenic peptides for GPCRs are useful in detecting
XX      an antibody against a particular GPCR, and in the production of specific
XX      antibodies. The peptides and antibodies are also useful for detecting the
XX      presence or absence of corresponding GPCRs. The antigenic peptides for
XX      GPCRs and antibodies are useful for diagnosing and designing drugs for
XX      treating immune-related diseases, growth-related diseases, cell
XX      regeneration-related disease, immunological-related cell proliferative
XX      diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
XX      atherosclerosis, bacterial, fungal, protozoan or viral infections,
XX      osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
XX      inflammation, allergies, Crohn's disease, diabetes, graft versus host
XX      disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
XX      anxiety, depression, schizophrenia, dementia, mental retardation, memory
XX      loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
XX      hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
XX      any other disorder in which GPCRs are involved. The antibodies may be
XX      used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode
XX      GPCR proteins given in ABP81675 to ABP82018, which are used in the
XX      exemplification of the present invention.
XX      Sequence 2923 AA:
XX      Query Match 100.0%; Score 15545; DB 24; Length 2923;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 MRSPATGVPPLPTPPPLLLLLLPPLLDGQVGPGRSGSGGACAPMGWLCPS 60
Db      1 MRSPATGVPPLPTPPPLLLLLLPPLLDGQVGPGRSGSGGACAPMGWLCPS 60
Qy      61 SASNLMYTRCRDAGTELGHVPHHDGLRWCPSEBAHILPAPPAEGGCPWSCRLIGIT 120

```

DB 61 SAKNLMYTSRCRDAETLGHLPVHHDGLRVCPESEAHILPBPAREGCPMSCLLIGIG 120  
QY 121 GHLSPOCKLTPREHPECLKAPRLRCSCKLAOPGLRAGERSPRESIGRRKNWTAPO 180  
DB 121 GHLSPOCKLTPREHPECLKAPRLRCSCKLAOPGLRAGERSPRESIGRRKNWTAPO 180  
QY 181 POPPSYQATVPENQPACTPVASJLRAIDPDEGEAGRLTYMDALFDSRSNOFSLDPVTGA 240  
DB 181 POPPSYQATVPENQPACTPVASJLRAIDPDEGEAGRLTYMDALFDSRSNOFSLDPVTGA 240  
QY 241 VTTAAEELDRCTKSTHVRVTAODHGMRRSALATLTLLVTDTMDHPVFEQOIKESLRE 300  
DB 241 VTTAAEELDRCTKSTHVRVTAODHGMRRSALATLTLLVTDTMDHPVFEQOIKESLRE 300  
QY 301 NLEGEVETLTVRATDGPAPNANILYRLLEGSGSPSEVEIDPRSGVIRTRBPVREBY 360  
DB 301 NLEGEVETLTVRATDGPAPNANILYRLLEGSGSPSEVEIDPRSGVIRTRBPVREBY 360  
QY 361 ESYQATVEASDQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVQVREDVTPGAPVLR 420  
DB 361 ESYQATVEASDQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVQVREDVTPGAPVLR 420  
QY 421 VTTASDDBDKGNAYVHYSIMGNAKGQFYLDQAGALDVVSPLDYETTKETTLVRAQDGG 480  
DB 421 VTTASDDBDKGNAYVHYSIMGNAKGQFYLDQAGALDVVSPLDYETTKETTLVRAQDGG 480  
QY 481 RPLPSNVSLGVTQVLDINDNAPIFVSTPROATVLSVPLGYVTLVHQAIDADAGNARL 540  
DB 481 RPLPSNVSLGVTQVLDINDNAPIFVSTPROATVLSVPLGYVTLVHQAIDADAGNARL 540  
QY 541 EYRLAGVGHDPPTTINNCGTMSVAAELDRBYDFVSGVEABDHGTALTAASVTVY 600  
DB 541 EYRLAGVGHDPPTTINNCGTMSVAAELDRBYDFVSGVEABDHGTALTAASVTVY 600  
QY 601 LDVNDNNPFTQOBYTYRLNEDAAVGTSVTVSAVNDASVTTYQITSENTNRSTIS 660  
DB 601 LDVNDNNPFTQOBYTYRLNEDAAVGTSVTVSAVNDASVTTYQITSENTNRSTIS 660  
QY 661 QSGGGLVSLALPDYKLERQYVLAVTASDGTRODTAQIVNVVDATHTHVPFQSHXTYN 720  
DB 661 QSGGGLVSLALPDYKLERQYVLAVTASDGTRODTAQIVNVVDATHTHVPFQSHXTYN 720  
QY 721 VNEBRPAGTTVILISATDEDTGENNARTTYMEDSIFQFRIDADTGAVTTOAELDYEDQYS 780  
DB 721 VNEBRPAGTTVILISATDEDTGENNARTTYMEDSIFQFRIDADTGAVTTOAELDYEDQYS 780  
QY 781 YTLAITARNDGIRQKSDTTYLLEILVNDVNDNAPOFLRDSYQGSVYEDVPPFYSVLQISAT 840  
DB 781 YTLAITARNDGIRQKSDTTYLLEILVNDVNDNAPOFLRDSYQGSVYEDVPPFYSVLQISAT 840  
QY 841 DRDSGLNGRVFTYFOGSDGDGDPIVESTSGIVRTLRLDRBNVAQYVLAAYVADKMP 900  
DB 841 DRDSGLNGRVFTYFOGSDGDGDPIVESTSGIVRTLRLDRBNVAQYVLAAYVADKMP 900  
QY 901 ARTPMEVTVTLVDVNNPVPFEODEPVEFVENSPIGLAVARTATDPDEGTNAQIMYOI 960  
DB 901 ARTPMEVTVTLVDVNNPVPFEODEPVEFVENSPIGLAVARTATDPDEGTNAQIMYOI 960  
QY 961 VEBNIEVPOLDIFSGELTALVDLDEDRBEYLVIOATSAPVSRATYHVLRLDNDNP 1020  
DB 961 VEBNIEVPOLDIFSGELTALVDLDEDRBEYLVIOATSAPVSRATYHVLRLDNDNP 1020  
QY 1021 PVLGNFEILFNNYVTKNSSPFGAIGRVPAHDPDISLSLTFSPERGNELSLVLNASTG 1080  
DB 1021 PVLGNFEILFNNYVTKNSSPFGAIGRVPAHDPDISLSLTFSPERGNELSLVLNASTG 1080  
QY 1081 EKLKSLALDNNRPLBAIMSVLSDGVHSTAOALATYTTTDMTLHSTTLRLJEDMSPER 1140  
DB 1081 EKLKSLALDNNRPLBAIMSVLSDGVHSTAOALATYTTTDMTLHSTTLRLJEDMSPER 1140  
QY 1141 FLSPPLGLFIQAAVATLATPDPHVVFVNVORDTDAAGHILNVSLSVGPPGGGAPPL 1200  
DB 1141 FLSPPLGLFIQAAVATLATPDPHVVFVNVORDTDAAGHILNVSLSVGPPGGGAPPL 1200

DB 1141 FLSPPLGLFIQAAVATLATPDPHVVFVNVORDTDAAGHILNVSLSVGPPGGGAPPL 1200  
QY 1201 PSEDLOERLYNRSILTLTISAORVLPDDNITCLAREPCENYMCVSLRPSAPLASS 1260  
DB 1201 PSEDLOERLYNRSILTLTISAORVLPDDNITCLAREPCENYMCVSLRPSAPLASS 1260  
QY 1261 VLFRLPHVGLRCHCPGFTGDYCEYEDVLCYSPPCGPHGRCSRREGYTLACBDGYG 1320  
DB 1261 VLFRLPHVGLRCHCPGFTGDYCEYEDVLCYSPPCGPHGRCSRREGYTLACBDGYG 1320  
QY 1321 EHCYVSAAGRCTPVCKNGTCVNLVGGFFCDPSPGDFEKPVCQVTRSPANSFTIF 1380  
DB 1321 EHCYVSAAGRCTPVCKNGTCVNLVGGFFCDPSPGDFEKPVCQVTRSPANSFTIF 1380  
QY 1381 RGLRORFHFTLTSATKBRDGLLYNRPFKKHDPVALLVLEVOEVOULTFSAGESTTYS 1440  
DB 1381 RGLRORFHFTLTSATKBRDGLLYNRPFKKHDPVALLVLEVOEVOULTFSAGESTTYS 1440  
QY 1441 PVPGVSDQWHTVQLKYNPRLIGOTGLPOGSEBQKVAVTVVDCDGTVALRFGSVLG 1500  
DB 1441 PVPGVSDQWHTVQLKYNPRLIGOTGLPOGSEBQKVAVTVVDCDGTVALRFGSVLG 1500  
QY 1501 NYSCAAQTOGGSKKSIDLTPGLLGGVBDLPESPPVNRQPVGCMRNLQVDSRHIDMAD 1560  
DB 1501 NYSCAAQTOGGSKKSIDLTPGLLGGVBDLPESPPVNRQPVGCMRNLQVDSRHIDMAD 1560  
QY 1561 FIANNGTVPCCAKKQVDCSNTCHNGTICVNOMDAPSCBCLGFGGKCAQOMANPOHFL 1620  
DB 1561 FIANNGTVPCCAKKQVDCSNTCHNGTICVNOMDAPSCBCLGFGGKCAQOMANPOHFL 1620  
QY 1621 GSSLVAMHGLSLPISQPMYLSLMPFTRQADGVLLQAITRGRSTITLQLEBGMVLSVGT 1680  
DB 1621 GSSLVAMHGLSLPISQPMYLSLMPFTRQADGVLLQAITRGRSTITLQLEBGMVLSVGT 1680  
QY 1681 GLQASLLEPRANDGMHHAQLALGASGGGHAILSPDVGQQAENLGRHLGHLIS 1740  
DB 1681 GLQASLLEPRANDGMHHAQLALGASGGGHAILSPDVGQQAENLGRHLGHLIS 1740  
QY 1741 NITVGGIIPGAGVARGRGLQGVRSVDTPEGVNSLDPSSHESINVEQCSLPDPCSN 1800  
DB 1741 NITVGGIIPGAGVARGRGLQGVRSVDTPEGVNSLDPSSHESINVEQCSLPDPCSN 1800  
QY 1801 PCPANSYCSNMDYSYSCDPEYGDNCTNVCNLPCEHOSYCTRKPAPPHGYTCBPPN 1860  
DB 1801 PCPANSYCSNMDYSYSCDPEYGDNCTNVCNLPCEHOSYCTRKPAPPHGYTCBPPN 1860  
QY 1861 YLGPYCETRLIDPCCRGMMGHPTGCPNCNCDVSKGFPDPCNKTSGRCHCKENHYRPGSPT 1920  
DB 1861 YLGPYCETRLIDPCCRGMMGHPTGCPNCNCDVSKGFPDPCNKTSGRCHCKENHYRPGSPT 1920  
QY 1921 CLLCDYPTGSLSRVCDPBDQCPCKPGVIGRQCDRCNPFALVYTNNGCEVYDSCPAI 1980  
DB 1921 CLLCDYPTGSLSRVCDPBDQCPCKPGVIGRQCDRCNPFALVYTNNGCEVYDSCPAI 1980  
QY 1981 EAGIWPPTTRFGLPAAPCPYKSGFSTAVRHCDBERGMPLPMLFNCTSTTFSELKFAERL 2040  
DB 1981 EAGIWPPTTRFGLPAAPCPYKSGFSTAVRHCDBERGMPLPMLFNCTSTTFSELKFAERL 2040  
QY 2041 ORNESGLSGRSQOALILRNATHTAGYFGSDVVAQAOLTRLLAHSTORGFLSATO 2100  
DB 2041 ORNESGLSGRSQOALILRNATHTAGYFGSDVVAQAOLTRLLAHSTORGFLSATO 2100  
QY 2101 DVHFTENLLRVGSALDLTANKRHWELIQTOSTGATMLLOHYEAYASALAQNNRHITYLSPF 2160  
DB 2101 DVHFTENLLRVGSALDLTANKRHWELIQTOSTGATMLLOHYEAYASALAQNNRHITYLSPF 2160  
QY 2161 TTVTNNIYISVVRIDKGNPAKAPLRYEALRGEOPDLLETVIILPESVPRRPPVVRAG 2220  
DB 2161 TTVTNNIYISVVRIDKGNPAKAPLRYEALRGEOPDLLETVIILPESVPRRPPVVRAG 2220  
QY 2221 PGEAOPEELARORRHPELSOGSEAVASVIYRITLAGLLPHNYDDDKSLRLVKKPINT 2280  
DB 2221 PGEAOPEELARORRHPELSOGSEAVASVIYRITLAGLLPHNYDDDKSLRLVKKPINT 2280

```

QY 2281 PVSISVHDEDELLPRLADKPVTVQRLLETERTIPICVFNHNSILVSGTGMARGGE 2340
DB 2281 PVSISVHDEDELLPRLADKPVTVQRLLETERTIPICVFNHNSILVSGTGMARGGE 2340
QY 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLVYVALGVTLAALLTTFEFLU 2400
DB 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLVYVALGVTLAALLTTFEFLU 2400
QY 2401 LRLRNSNGIRRMNLTALGLAQLVFLGLINQADLPACTVIAILLFPLVCTFSWALIE 2460
DB 2401 LRLRNSNGIRRMNLTALGLAQLVFLGLINQADLPACTVIAILLFPLVCTFSWALIE 2460
QY 2461 ALHLVYALTEVADVMTGPMRFYTMKGWCPAFITGLAVGLDPGYGNPFCWLSTYDTLI 2520
DB 2461 ALHLVYALTEVADVMTGPMRFYTMKGWCPAFITGLAVGLDPGYGNPFCWLSTYDTLI 2520
QY 2521 WSPFAGVAVAVMSVFLYTLAARASCAORQGEKKGPVSGLOPSPAVILLLSATWLLAL 2580
DB 2521 WSPFAGVAVAVMSVFLYTLAARASCAORQGEKKGPVSGLOPSPAVILLLSATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFATCNCIQGPFIPLSYVVLKSKVRKALKACSRKPSPPDALTTSKTL 2640
DB 2581 LSVNSDTLLFHYLFATCNCIQGPFIPLSYVVLKSKVRKALKACSRKPSPPDALTTSKTL 2640
QY 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTSRGSKSQPSYIPLAREESALNPGQGPGLG 2700
DB 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTSRGSKSQPSYIPLAREESALNPGQGPGLG 2700
QY 2701 DGSLPLFEGQDOHDPTDSDSLSLDDQSGSYASTHSDSEEEEEEEREAAPGREG 2760
DB 2701 DGSLPLFEGQDOHDPTDSDSLSLDDQSGSYASTHSDSEEEEEEEREAAPGREG 2760
QY 2761 WDSLLGPGAERLPLHSTPKDGGPGPKAPMGDPGTAKESSGNGAPBEERLENGDALSR 2820
DB 2761 WDSLLGPGAERLPLHSTPKDGGPGPKAPMGDPGTAKESSGNGAPBEERLENGDALSR 2820
QY 2821 ESSLGRLPSSSAQPHKGLIKKCLPTISKSSILRLPLEQCTGSSRGSSASGSRGGPPP 2880
DB 2821 ESSLGRLPSSSAQPHKGLIKKCLPTISKSSILRLPLEQCTGSSRGSSASGSRGGPPP 2880
QY 2881 RPPRQSLQEQNLGWPFIAMSIXAGTVDEDSGSEFLPFNFLLH 2923
DB 2881 RPPRQSLQEQNLGWPFIAMSIXAGTVDEDSGSEFLPFNFLLH 2923

```

RESULT 4  
AAU07053  
ID AAU07053 standard; Protein; 2956 AA.

```

XX 24-OCT-2001 (fixed entry)
XX AAU07053;
XX
XX Human Flamingo polypeptide.
XX
XX Flamingo; human; G-protein coupled receptor; diabetes; protozoan; cancer;
XX signal transduction pathway; bacterial; fungal; viral; anorexia; bulimia;
XX asthma; Parkinson's disease; acute heart failure; HIV; obesity; anxiety;
XX hypertension; hypertension; urinary retention; angina pectoris; allergy;
XX myocardial infarction; stroke; ulcer; schizophrenia; psychotic disorder;
XX benign prostatic hypertrophy; neurological disorder; manic depression;
XX delirium; dementia; severe mental retardation; Huntington's disease;
XX Gilles de la Tourette's syndrome; antibacterial; antifungal; antiviral;
XX antiprotocozal; anti-HIV; anorectic; antiasthmatic; antiparkinsonian;
XX cardiatic; cerebroprotective; neuroprotective; antidepressant;
XX anticonvulsant; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200161003-A1.
XX
XX 23-AUG-2001.

```

```

XX 19-FEB-2001; 2001WO-GB00680.
XX
XX 19-FEB-2000; 2000GB-0004196.
XX
XX (SMK ) SMITHKLINE BEECHAM PLC.
XX
XX Testa TT;
XX
XX WPI; 2001-502792/55.
XX
XX N-PSDB; AAS11677.
XX
XX An isolated Flamingo polypeptide useful for treating diseases such as
XX HIV, cancer, asthma, Parkinson's disease, acute heart failure,
XX osteoporosis -
XX
XX Claim 2; Page 28-29; 66pp; English.
XX
XX The sequence represents a human Flamingo polypeptide. Flamingo is a
XX member of the G-protein coupled receptor family, which is involved in
XX signal transduction pathways. By screening to identify compounds that
XX stimulate or inhibit the function or level of the protein, treatments can
XX be developed for various diseases and bacterial, fungal, protozoan and
XX viral infections, including HIV, cancer, diabetes, obesity, anorexia,
XX bulimia, asthma, Parkinson's disease, acute heart failure, hypertension,
XX stroke, ulcers, allergies and benign prostatic hypertrophy. Also
XX treatable are psychotic and neurological disorders such as anxiety,
XX schizophrenia, manic depression, delirium, dementia, severe mental
XX retardation, Huntington's disease and Gilles de la Tourette's syndrome.
XX
XX Sequence 2956 AA;
XX
XX Query Match 99.8%; Score 15518.5; DB 22; Length 2956;
XX Best Local Similarity 98.9%; Pred. No. 0;
XX Matches 2923; Conservative 0; Mismatches 0; Indels 33; Gaps 1;
XX
XX 1 MRSPATGVPLTPPPPLLLLLLLLPPLLDQVPCSLGSRGSSGACAPMGWLCPS 60
XX 1 MRSPATGVPLTPPPPLLLLLLLLPPLLDQVPCSLGSRGSSGACAPMGWLCPS 60
XX
XX 61 SASNLMLYTSRCRAGTTLTHLVPHDGLRWCPESBAHPLPAPPGCWCSCRLGIG 120
XX 61 SASNLMLYTSRCRAGTTLTHLVPHDGLRWCPESBAHPLPAPPGCWCSCRLGIG 120
XX
XX 61 SASNLMLYTSRCRAGTTLTHLVPHDGLRWCPESBAHPLPAPPGCWCSCRLGIG 120
XX 61 SASNLMLYTSRCRAGTTLTHLVPHDGLRWCPESBAHPLPAPPGCWCSCRLGIG 120
XX
XX 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRAGERSPRESLGRRKRNVTAPQ 180
XX 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRAGERSPRESLGRRKRNVTAPQ 180
XX
XX 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRAGERSPRESLGRRKRNVTAPQ 180
XX 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRAGERSPRESLGRRKRNVTAPQ 180
XX
XX 181 FQPPSYQATVBNOPAGTPVVALRAIDPDEGEAGRLLEYMDALFDSRNSQFSLDPVTGA 240
XX 181 FQPPSYQATVBNOPAGTPVVALRAIDPDEGEAGRLLEYMDALFDSRNSQFSLDPVTGA 240
XX
XX 241 VTTABEIDRETKSTHVRVTAQDHGMPRRSALATLTLLVTNDHDVFEQEKESLRE 300
XX 241 VTTABEIDRETKSTHVRVTAQDHGMPRRSALATLTLLVTNDHDVFEQEKESLRE 300
XX
XX 241 VTTABEIDRETKSTHVRVTAQDHGMPRRSALATLTLLVTNDHDVFEQEKESLRE 300
XX 241 VTTABEIDRETKSTHVRVTAQDHGMPRRSALATLTLLVTNDHDVFEQEKESLRE 300
XX
XX 301 NLEFGEVFLVTRADGAPPAANILVRLLEGSGSPSRVPIIDRSQVIRRGVNDREB 360
XX 301 NLEFGEVFLVTRADGAPPAANILVRLLEGSGSPSRVPIIDRSQVIRRGVNDREB 360
XX
XX 301 NLEFGEVFLVTRADGAPPAANILVRLLEGSGSPSRVPIIDRSQVIRRGVNDREB 360
XX 301 NLEFGEVFLVTRADGAPPAANILVRLLEGSGSPSRVPIIDRSQVIRRGVNDREB 360
XX
XX 361 ESYQLVTEASDQDGPGRSTTAAVFLSVEDNDNAPQFSEKRYVVOVREDVTPGAPYLR 420
XX 361 ESYQLVTEASDQDGPGRSTTAAVFLSVEDNDNAPQFSEKRYVVOVREDVTPGAPYLR 420
XX
XX 421 VTTADBRKGSNAVHYHISMGNGAGGYLDAGTALVVPGLDYETTKYTLRRAQDGG 480
XX 421 VTTADBRKGSNAVHYHISMGNGAGGYLDAGTALVVPGLDYETTKYTLRRAQDGG 480
XX
XX 421 VTTADBRKGSNAVHYHISMGNGAGGYLDAGTALVVPGLDYETTKYTLRRAQDGG 480
XX 421 VTTADBRKGSNAVHYHISMGNGAGGYLDAGTALVVPGLDYETTKYTLRRAQDGG 480
XX
XX 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVJESVPGYLVATFOADADAGDNARL 540
XX 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVJESVPGYLVATFOADADAGDNARL 540
XX
XX 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVJESVPGYLVATFOADADAGDNARL 540
XX 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVJESVPGYLVATFOADADAGDNARL 540

```

QY 541 EYRLAGVHDPEFTINNGTGMISVAELREBYDYSFGVEARDHGTALITASASVTV 600  
 DB 541 EYRLAGVHDPEFTINNGTGMISVAELREBYDYSFGVEARDHGTALITASASVTV 600  
 QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYTVSAVDRDAHSVTVYITIGANTRNRSITS 660  
 DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYTVSAVDRDAHSVTVYITIGANTRNRSITS 660  
 QY 661 QSGGGVSLALPDYDYLEROYVLAATASDGTRODTAQIVNVTADATNHPVFOSSHXTN 720  
 DB 661 QSGGGVSLALPDYDYLEROYVLAATASDGTRODTAQIVNVTADATNHPVFOSSHXTN 720  
 QY 721 VNEDRAGTTVVLISATDEBTGENARITYPMEDSIPQFRIDATGAVTTOABLDYEDYS 780  
 DB 721 VNEDRAGTTVVLISATDEBTGENARITYPMEDSIPQFRIDATGAVTTOABLDYEDYS 780  
 QY 781 YTLATIRARNGIPQKSDTTYLETUNVDNMDAQTALRDSQGSYEDVPPFSVLOISAT 840  
 DB 781 YTLATIRARNGIPQKSDTTYLETUNVDNMDAQTALRDSQGSYEDVPPFSVLOISAT 840  
 QY 841 DRDSGLNGRVFTYFOGDDGDDGFIVESTSGIVRTLRRLDRNVAAQYVLAAYDKMP 900  
 DB 841 DRDSGLNGRVFTYFOGDDGDDGFIVESTSGIVRTLRRLDRNVAAQYVLAAYDKMP 900  
 QY 901 ARTPEVTVTVLDVNDNPPVFEDEDFVVEENSPIGLAVARVATDDEGTNAQIMYQI 960  
 DB 901 ARTPEVTVTVLDVNDNPPVFEDEDFVVEENSPIGLAVARVATDDEGTNAQIMYQI 960  
 QY 961 VEGNTPEVROLOIFSGELTALVDYEDRBEVYVLOATSAPLYSAATNVLDRNDP 1020  
 DB 961 VEGNTPEVROLOIFSGELTALVDYEDRBEVYVLOATSAPLYSAATNVLDRNDP 1020  
 QY 1021 PVLNGEILFNNVYTRRSSSPFGAIGRPAHDPDISLSITYSFERGNEISLVTLNASTG 1080  
 DB 1021 PVLNGEILFNNVYTRRSSSPFGAIGRPAHDPDISLSITYSFERGNEISLVTLNASTG 1080  
 QY 1081 ELKLSALDNNRPLEAIVSVSDGVHSTVAOCALVTITIDEMTLHSTITLREDSMPER 1140  
 DB 1081 ELKLSALDNNRPLEAIVSVSDGVHSTVAOCALVTITIDEMTLHSTITLREDSMPER 1140  
 QY 1141 FLSPLGLFIOAAVATLAPPHDVVFNVOPTDAGGHIINVSLSVGPFGGSGPPL 1200  
 DB 1141 FLSPLGLFIOAAVATLAPPHDVVFNVOPTDAGGHIINVSLSVGPFGGSGPPL 1200  
 QY 1201 PSEDOERLYLNSLTALISACVRLPEDDNICLRBCEMYMRCVSLARPDSSAPFIASS 1260  
 DB 1201 PSEDOERLYLNSLTALISACVRLPEDDNICLRBCEMYMRCVSLARPDSSAPFIASS 1260  
 QY 1261 VLFRLPHVVGJLRCPGFTGDCYCTEVDLYSRPCGPHGRCSRREGGYTCLCRDGYTG 1320  
 DB 1261 VLFRLPHVVGJLRCPGFTGDCYCTEVDLYSRPCGPHGRCSRREGGYTCLCRDGYTG 1320  
 QY 1321 EHCEVASRSGRCRGVCKNGGTGCVNLJNGFGKDCSGDPFEKXYCOVTRSPRHSFITF 1380  
 DB 1321 EHCEVASRSGRCRGVCKNGGTGCVNLJNGFGKDCSGDPFEKXYCOVTRSPRHSFITF 1380  
 QY 1381 RGLRQRFHFTALSPATKERDGLLYNGRFXEKHDFVLALEVOEVOULFSGESTTYS 1440  
 DB 1381 RGLRQRFHFTALSPATKERDGLLYNGRFXEKHDFVLALEVOEVOULFSGESTTYS 1440  
 QY 1441 PVPFGVSDGQMTVOLKYNKPLLGOTGLPOGSRQKAVAVTVDGCCTGVALRFGSVTG 1500  
 DB 1441 PVPFGVSDGQMTVOLKYNKPLLGOTGLPOGSRQKAVAVTVDGCCTGVALRFGSVTG 1500  
 QY 1501 NYSCAAQGTGGSKSLDLTGPLLGVPDLPSFPBAMQFVGCMBRNOVDSRHDMD 1560  
 DB 1501 NYSCAAQGTGGSKSLDLTGPLLGVPDLPSFPBAMQFVGCMBRNOVDSRHDMD 1560  
 QY 1561 FIANNQTPGCPAKXVCDSNTCHNGTCVNWDAFSCCEPLAGFGSKCAQEMANQHFL 1620  
 DB 1561 FIANNQTPGCPAKXVCDSNTCHNGTCVNWDAFSCCEPLAGFGSKCAQEMANQHFL 1620  
 QY 1621 GSSLVAMHGLSLPISQPWYLSLMPRTROADGVLLQATIRGRSTITLQALREGHVMSVSGT 1680

DB 1621 GSSLVAMHGLSLPISQPWYLSLMPRTROADGVLLQATIRGRSTITLQALREGHVMSVSGT 1680  
 QY 1681 GLQASSLALREGRANDGMHHAQALASGSGPCHAIIISFDYGOQABENLPRHLGLHS 1740  
 DB 1681 GLQASSLALREGRANDGMHHAQALASGSGPCHAIIISFDYGOQABENLPRHLGLHS 1740  
 QY 1741 NITVGGIPGAPAGVARGRGCLQGVRSVDTPEGVNSLDPFSHGESINVEQCSLPDPCSN 1800  
 DB 1741 NITVGGIPGAPAGVARGRGCLQGVRSVDTPEGVNSLDPFSHGESINVEQCSLPDPCSN 1800  
 QY 1801 PCPANSYCSNDWDSYSCSDPGYGDNCTNVCDINPCESQSVCTRKPSAPHGYTCBP 1860  
 DB 1801 PCPANSYCSNDWDSYSCSDPGYGDNCTNVCDINPCESQSVCTRKPSAPHGYTCBP 1860  
 QY 1861 YLGPYCEFRIDOPCRGMMGHPTGCPKCDVSKGDPUCNKTSGCHCKENHYRPGSP 1920  
 DB 1861 YLGPYCEFRIDOPCRGMMGHPTGCPKCDVSKGDPUCNKTSGCHCKENHYRPGSP 1920  
 QY 1921 CLICDCYPTGSLSRVCDPEBQCPCKPGVIGRQCDRCNDPFAEYTNNGEYVYDSCPPAI 1980  
 DB 1921 CLICDCYPTGSLSRVCDPEBQCPCKPGVIGRQCDRCNDPFAEYTNNGEYVYDSCPPAI 1980  
 QY 1981 EAGIWPPTRGFLPAAPCPKGSFGTAVRCHDERGMLPMLFNCTSTFSELKGFARL 2040  
 DB 1981 EAGIWPPTRGFLPAAPCPKGSFGTAVRCHDERGMLPMLFNCTSTFSELKGFARL 2040  
 QY 2041 QNBSGLDSGRSQQALALIRNAHTAGYFGSDVKVAQOLATRLAHSTORGJLSATQ 2100  
 DB 2041 QNBSGLDSGRSQQALALIRNAHTAGYFGSDVKVAQOLATRLAHSTORGJLSATQ 2100  
 QY 2101 DVHFTENLIRVGSALDPTANKRHELIQOTEGTAMLLQHYEAASALAQNMHTYLSPF 2160  
 DB 2101 DVHFTENLIRVGSALDPTANKRHELIQOTEGTAMLLQHYEAASALAQNMHTYLSPF 2160  
 QY 2161 TIVPNTIVISVRLDKGNPAKAPRYBALRGEOPDLETTVILPESYFRETPPVVRAG 2220  
 DB 2161 TIVPNTIVISVRLDKGNPAKAPRYBALRGEOPDLETTVILPESYFRETPPVVRAG 2220  
 QY 2221 PGEAOPEELARORRHELSQGEAVASVITYRTLAGLLPHNYOPDKSLVVPKPIINT 2280  
 DB 2221 PGEAOPEELARORRHELSQGEAVASVITYRTLAGLLPHNYOPDKSLVVPKPIINT 2280  
 QY 2281 PVSISVHDBELLPRALDKPEVTVOFLLSTBERKPICVFNHNSILVSGTGSARCE 2340  
 DB 2281 PVSISVHDBELLPRALDKPEVTVOFLLSTBERKPICVFNHNSILVSGTGSARCE 2340  
 QY 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGEILPKLTLYVALGVTLAALLTFEFLT 2400  
 DB 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGEILPKLTLYVALGVTLAALLTFEFLT 2400  
 QY 2401 LRIIRSNQHGIRRNLTALAGLAQVFLGIGNQADLPACTVIALIHLFLYCTFSWALLE 2460  
 DB 2401 LRIIRSNQHGIRRNLTALAGLAQVFLGIGNQADLPACTVIALIHLFLYCTFSWALLE 2460  
 QY 2461 ALHLRYALTBYRDVNTGEMRFYMLGMGVPAFITGLAVGLDPEGYNDPCWLSYDTLI 2520  
 DB 2461 ALHLRYALTBYRDVNTGEMRFYMLGMGVPAFITGLAVGLDPEGYNDPCWLSYDTLI 2520  
 QY 2521 WSPAGPVAFAVMSVFLYTLAARASCAAQROGFEKKGVSGLQSPFALLLSAWMLAL 2580  
 DB 2521 WSPAGPVAFAVMSVFLYTLAARASCAAQROGFEKKGVSGLQSPFALLLSAWMLAL 2580  
 QY 2581 LSVNBDTLFHYLFATCNCIOGPFIFLSYVLSKEVRALKLACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNBDTLFHYLFATCNCIOGPFIFLSYVLSKEVRALKLACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNCSPYADGRLYQPYGDSAGLSHSTRSGKSQSPYIFPLLRBEBALMPGOQPGIG 2700  
 DB 2641 TSSYNCSPYADGRLYQPYGDSAGLSHSTRSGKSQSPYIFPLLRBEBALMPGOQPGIG 2700  
 QY 2701 DPGSLFLEGDQOQHDPDUDSDSLBDOSGSYASTSSDSBEBEBEBEAAAPRGOG 2760



DB 2701 DRGSLFLEGGDQHDPDNDSDLSLEDDQSGSYASTHSDSEBEEBEEBAPPGGQ 2760  
 QY 2761 WDLILGPGAEIRLPLHSTPDGPGPGKAPGDPFGTTAESSGNGAPERHLRENGDALSR 2820  
 DB 2761 WDLILGPGAEIRLPLHSTPDGPGPGKAPGDPFGTTAESSGNGAPERHLRENGDALSR 2820  
 QY 2821 ESSLGPELPSSAQPHK-----GILKKCLPTI 2847  
 DB 2821 ESSLGPELPSSAQPHKEMGTGTPSCRAVLVSLIPIILMPLHTAPPRPTGIKKCLPTI 2880  
 QY 2848 SEKSSILRLPLEOCTGSSRGSSASBEGSRGPPPPPPROSLQEOINQVWPIAMSIKATV 2907  
 DB 2881 SEKSSILRLPLEOCTGSSRGSSASBEGSRGPPPPPPROSLQEOINQVWPIAMSIKATV 2940  
 QY 2908 DEDSSGSEPLFFNFTH 2923  
 DB 2941 DEDSSGSEPLFFNFTH 2956

RESULT 5  
 ABU11556  
 ID ABU11556 standard; Protein; 2894 AA.  
 AC ABU11556;  
 XX  
 DT 12-FEB-2003 (first entry)  
 DE Human MDDT polypeptide SEQ ID 503.  
 XX  
 MDPT; human; disease detection and treatment molecule polypeptide;  
 anti-inflammatory; immunosuppressive; osteopathic; cyclostatic; anti-HIV;  
 haemostatic; nephrotropic; antianemic; antipsoriatic; hepatotropic;  
 gene therapy; protein replacement therapy; cell proliferative disorder;  
 cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;  
 anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 psoriasis; hepatitis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PE 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 29-MAR-2001; 2001US-280068P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 17-MAY-2001; 2001US-291849P.  
 PR 19-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J,  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;  
 PI Daugherty SC, Dam TC, Liu TP, Nguyen DA, Kleeefeld Y, Gerstein EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,  
 PI Flores V, Marwala R, Lo A, Lan RY, Urashka ME;  
 XX  
 WIPI; 2003-058431/05.  
 DR N-PEDB; ABX34546.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukaemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis  
 XX  
 PS Claim 27; SEQ ID NO 503; 339pp + Sequence Listing; English.

XX This invention describes a novel disease detection and treatment molecule  
 CC polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,  
 CC osteopathic, cyclostatic, anti-HIV, haemostatic, nephrotropic,  
 CC antianemic, antipsoriatic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or  
 CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded  
 CC by ABU11450-ABU11845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_ptc\_sequences.  
 CC  
 XX

Sequence 2894 AA;

Query Match 99.0%; Score 15393; DB 24; Length 2894;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 LGDQVGPCRSLSGSRGSSGACAPMGWLCPSSASNTMLYTSRCRDAGTELTHGLVPHHDG 89  
 DB 1 LGDQVGPCRSLSGSRGSSGACAPMGWLCPSSASNTMLYTSRCRDAGTELTHGLVPHHDG 60  
 QY 90 LRWVCPSEBARIPLPPAPGEGPCRLIGTGHSPQCKTLPREHCLKAPRLRCOSCK 149  
 DB 61 LRWVCPSEBARIPLPPAPGEGPCRLIGTGHSPQCKTLPREHCLKAPRLRCOSCK 120  
 QY 150 LAQAPGLRAGRSRSEBSIGRRKRNVTAPQFQPSYQATYBKNOPATPARSARIDPD 209  
 DB 121 LAQAPGLRAGRSRSEBSIGRRKRNVTAPQFQPSYQATYBKNOPATPARSARIDPD 180  
 QY 210 EGBAGRLTYMDALFDRSNQPFSLDPVTGAVTAAEIDRETGSTHYRVTAQDHGERR 269  
 DB 181 EGBAGRLTYMDALFDRSNQPFSLDPVTGAVTAAEIDRETGSTHYRVTAQDHGERR 240  
 QY 270 SALATLTILVTNDMDHPVPEQCKESLRNLEVGREYLVRAITDGDAPPANILYRL 329  
 DB 241 SALATLTILVTNDMDHPVPEQCKESLRNLEVGREYLVRAITDGDAPPANILYRL 300  
 QY 330 ESGGSGSEVEFEIDPRSGVIRTRGPVDRBEVESYQLTVASDQGRDPGPRSTAAVFLSV 389  
 DB 301 ESGGSGSEVEFEIDPRSGVIRTRGPVDRBEVESYQLTVASDQGRDPGPRSTAAVFLSV 360  
 QY 390 EDDNDNAPQSEKRYVQVREDDVTGAPVLVITASDRDKGSNAVHYSIMSNGARQFYL 449  
 DB 361 EDDNDNAPQSEKRYVQVREDDVTGAPVLVITASDRDKGSNAVHYSIMSNGARQFYL 420  
 QY 450 DAQFGALDVNSPDIYETTKKTYTLARAQDQGRPPLSVSGLYTVQVLDINDNAIFPSTP 509  
 DB 421 DAQFGALDVNSPDIYETTKKTYTLARAQDQGRPPLSVSGLYTVQVLDINDNAIFPSTP 480  
 QY 510 FOATVLESVPAGYVLAFOAIDADAGNARLEVLAVGADPPTINGGWTISVAELD 569  
 DB 481 FOATVLESVPAGYVLAFOAIDADAGNARLEVLAVGADPPTINGGWTISVAELD 540  
 QY 570 REEVDFTSFGVEARDHGTPLATASASVTVLDVNDNPTFTQPEYTVRLNEDAAVGTSTV 629  
 DB 541 REEVDFTSFGVEARDHGTPLATASASVTVLDVNDNPTFTQPEYTVRLNEDAAVGTSTV 600  
 QY 630 VTVSAVDRDAHSVTVYQITSGNTRNRPSTISQSGGIVSALPLDYLERQYVLAATASD 689  
 DB 601 VTVSAVDRDAHSVTVYQITSGNTRNRPSTISQSGGIVSALPLDYLERQYVLAATASD 660  
 QY 690 GTRQDTAQIVVNTDANTHRPVPQSSHVTYVNVNEDRAGTGVVLISATDDEGNARITY 749  
 DB 661 GTRQDTAQIVVNTDANTHRPVPQSSHVTYVNVNEDRAGTGVVLISATDDEGNARITY 720

QY 750 FMEBDFIOPFIADDTGAVTTQAEILDYEDQVSYTLATARDNGIPIQKSDTTYLEILVNDVN 809  
 DB 721 FMEBDFIOPFIADDTGAVTTQAEILDYEDQVSYTLATARDNGIPIQKSDTTYLEILVNDVN 780  
 QY 810 DNAPQFRLDSYQOSVVEEDVPFPTSVLQISATDSDSGNGRVEFTFGQDGDGDDDFIYEST 869  
 DB 781 DNAPQFRLDSYQOSVVEEDVPFPTSVLQISATDSDSGNGRVEFTFGQDGDGDDDFIYEST 840  
 QY 870 SGIVTRLRLRLDRBENVAQVYLRAVAVDKGMPARTPMEVTVTVLVNDNPNPVFEQDEDFV 929  
 DB 841 SGIVTRLRLRLDRBENVAQVYLRAVAVDKGMPARTPMEVTVTVLVNDNPNPVFEQDEDFV 900  
 QY 930 VEENSPGLAVATATDPDBGTNAQIMQIVBGNIPFVQOLDIFSGELTALVDLDEDR 989  
 DB 901 VEENSPGLAVATATDPDBGTNAQIMQIVBGNIPFVQOLDIFSGELTALVDLDEDR 960  
 QY 990 PEYVLVQATSAFLVSAATVHRLLDNDNPVYLGNEILLFNNVYNNRSSPFGAIGRY 1049  
 DB 961 PEYVLVQATSAFLVSAATVHRLLDNDNPVYLGNEILLFNNVYNNRSSPFGAIGRY 1020  
 QY 1050 PAHDPOISDLSYTSFERGNEILSVTLNASTGBELKLSBALDNNRPLBAIMSVLVSDGVHAY 1109  
 DB 1021 PAHDPOISDLSYTSFERGNEILSVTLNASTGBELKLSBALDNNRPLBAIMSVLVSDGVHAY 1080  
 QY 1110 TAOCALAVTITITDBMLTHSITTLRLMEDSPPERFISPLIGLFIQAVAAATLAPPHVFN 1169  
 DB 1081 TAOCALAVTITITDBMLTHSITTLRLMEDSPPERFISPLIGLFIQAVAAATLAPPHVFN 1140  
 QY 1170 QROTDAPAGHILNVLNLSVSGPPGPGGPPFLPSEDLQERLYNRSLLTATSAORVLPEDD 1229  
 DB 1141 QROTDAPAGHILNVLNLSVSGPPGPGGPPFLPSEDLQERLYNRSLLTATSAORVLPEDD 1200  
 QY 1230 NICLRBECENYMCVSVLAFDSSAPFIASSSVLERPIHPVGLRCRCRCPGPTGDCYETEV 1289  
 DB 1201 NICLRBECENYMCVSVLAFDSSAPFIASSSVLERPIHPVGLRCRCRCPGPTGDCYETEV 1260  
 QY 1290 DLYCYRCCGHHGCRSREGGTYCLCRDGYTGENCEVARSRCCTPGYCKNGCTVNLVNG 1349  
 DB 1261 DLYCYRCCGHHGCRSREGGTYCLCRDGYTGENCEVARSRCCTPGYCKNGCTVNLVNG 1320  
 QY 1350 GFKDCDPSGDFEKPXCOVTTTRSPPAHSFITFRGLRQRFHFLALSFATKRGDGLLYNGR 1409  
 DB 1321 GFKDCDPSGDFEKPXCOVTTTRSPPAHSFITFRGLRQRFHFLALSFATKRGDGLLYNGR 1380  
 QY 1410 FNEKHPFVALEVIQEOVOLTFSAGESTTVSPVPGSVSDQWHTVQLKTYNRPILGQTS 1469  
 DB 1381 FNEKHPFVALEVIQEOVOLTFSAGESTTVSPVPGSVSDQWHTVQLKTYNRPILGQTS 1440  
 QY 1470 LPOGSPROKAVVTVNDCDGVALARFGSVLGNYSCAAQGTQGSKSLDTGPILLGQVP 1529  
 DB 1441 LPOGSPROKAVVTVNDCDGVALARFGSVLGNYSCAAQGTQGSKSLDTGPILLGQVP 1500  
 QY 1530 DLBESFVNRAROFVGCARNIQVDSRHIDMADFLANNGTGVCPAKKNVCSNTCHNGGTC 1589  
 DB 1501 DLBESFVNRAROFVGCARNIQVDSRHIDMADFLANNGTGVCPAKKNVCSNTCHNGGTC 1560  
 QY 1590 VNOMDAFSCCEPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPMYLSIMFTRQA 1649  
 DB 1561 VNOMDAFSCCEPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPMYLSIMFTRQA 1620  
 QY 1650 DGVLLQAITRGSRITITLQREBHVMLSVBGTGLOASLRLBPRANDGMHNOALGAS 1709  
 DB 1621 DGVLLQAITRGSRITITLQREBHVMLSVBGTGLOASLRLBPRANDGMHNOALGAS 1680  
 QY 1710 GGEFHALLSFYDQOARAGNLAGRLHGLHLSNITVGGIPGPAGVANGFRGCLQGVAVSD 1769  
 DB 1681 GGEFHALLSFYDQOARAGNLAGRLHGLHLSNITVGGIPGPAGVANGFRGCLQGVAVSD 1740  
 QY 1770 TPBGSVNLDPHSHGASINVEOGCSLPDPCDSNPPCANYSICNDWDSYSCSDPGYTGDNCT 1829  
 DB 1741 TPBGSVNLDPHSHGASINVEOGCSLPDPCDSNPPCANYSICNDWDSYSCSDPGYTGDNCT 1800  
 QY 1830 NVCDLNPCEHQSVCTRKSPAPHGYTCECPNVLGYPCETRIDQCPRGWGMHPGTCPCNC 1889

DB 1801 NVCDLNPCEHQSVCTRKSPAPHGYTCECPNVLGYPCETRIDQCPRGWGMHPGTCPCNC 1860  
 QY 1890 DYSKGDPPCNTSGCHCKENHVRPPGSPCLLDCDQVYTGSLSRVCEDEDDQCPCKPBY 1949  
 DB 1861 DYSKGDPPCNTSGCHCKENHVRPPGSPCLLDCDQVYTGSLSRVCEDEDDQCPCKPBY 1920  
 QY 1950 IGRQCDRCNPPAEVYTNNGCEVYNDSCPRALBAGIMWPRTRGFLPAAABCPKSPGTAAR 2009  
 DB 1921 IGRQCDRCNPPAEVYTNNGCEVYNDSCPRALBAGIMWPRTRGFLPAAABCPKSPGTAAR 1980  
 QY 2010 HCDERHGMPLPULFNCTSTTSEBLKGAERLORNSGDLDSRSQOLALLRNATQHTAGY 2069  
 DB 1981 HCDERHGMPLPULFNCTSTTSEBLKGAERLORNSGDLDSRSQOLALLRNATQHTAGY 2040  
 QY 2070 FGSDDVAVAYQALTRLLAHSTQRGFLSATODVHFENMLARVGSALLDPANRHMWELIOO 2129  
 DB 2041 FGSDDVAVAYQALTRLLAHSTQRGFLSATODVHFENMLARVGSALLDPANRHMWELIOO 2100  
 QY 2130 TBGGTAMLLQHYEAVASALAQNMRRHTYLSPTITVTPNIVISYVRLDKGNFAGAKLPRYEA 2189  
 DB 2101 TBGGTAMLLQHYEAVASALAQNMRRHTYLSPTITVTPNIVISYVRLDKGNFAGAKLPRYEA 2160  
 QY 2190 LAGEQPPDLFTTVIIPBSYFRETPPVVRPAGGGAQOEPEBLARQORHPELSQGEAVASV 2249  
 DB 2161 LAGEQPPDLFTTVIIPBSYFRETPPVVRPAGGGAQOEPEBLARQORHPELSQGEAVASV 2220  
 QY 2250 IYRTLAGLLPHNYPDDKSLRVPKRPINTPVASISVHDEBELPRALDKVPTVQFRL 2309  
 DB 2221 IYRTLAGLLPHNYPDDKSLRVPKRPINTPVASISVHDEBELPRALDKVPTVQFRL 2280  
 QY 2310 ETEERTKPICVFMNHSILVSGTGSARCEVYFRNESHVSQCNHMTSPAVLMDVSRRE 2369  
 DB 2281 ETEERTKPICVFMNHSILVSGTGSARCEVYFRNESHVSQCNHMTSPAVLMDVSRRE 2340  
 QY 2370 NGBIILPLKLTLYVALGVTLAALLTPFRITLIRLSNQHGRNUTALGLAQLVFLIG 2429  
 DB 2341 NGBIILPLKLTLYVALGVTLAALLTPFRITLIRLSNQHGRNUTALGLAQLVFLIG 2400  
 QY 2430 INQADLPFACVYIALILHFLVCTFSMALLBAILHLYRALTEYRDVNTGPMREFTYMLGMGV 2489  
 DB 2401 INQADLPFACVYIALILHFLVCTFSMALLBAILHLYRALTEYRDVNTGPMREFTYMLGMGV 2460  
 QY 2490 PAFITGLAVGLDDEGYGNDFCMLSYDTLINSFAGPVAFAVMSVFLYTLAARASCAAQ 2549  
 DB 2461 PAFITGLAVGLDDEGYGNDFCMLSYDTLINSFAGPVAFAVMSVFLYTLAARASCAAQ 2520  
 QY 2550 ROGFEEKGVPVSGLOSPFALILLSTWILLALSUNSDTLFHYLPATCNCIOGPEIFLSY 2609  
 DB 2521 ROGFEEKGVPVSGLOSPFALILLSTWILLALSUNSDTLFHYLPATCNCIOGPEIFLSY 2580  
 QY 2610 VVLSKEVRKALKLACSRKSPDPALTTKSTLTSYNCSPYADGRLYQPYGDSAGLSHST 2669  
 DB 2581 VVLSKEVRKALKLACSRKSPDPALTTKSTLTSYNCSPYADGRLYQPYGDSAGLSHST 2640  
 QY 2670 SRSGKSQPSYIFPLLRBESSALNPGQPGGLGDPGLFLEGDQOQHDPTDSDSDLSLEBD 2729  
 DB 2641 SRSGKSQPSYIFPLLRBESSALNPGQPGGLGDPGLFLEGDQOQHDPTDSDSDLSLEBD 2700  
 QY 2730 QSGSYASTHSSDSEBEEBEEBEEAAPPQGGQMSLLGPABKRLPLHSTYTKDGGPGGKAP 2789  
 DB 2701 QSGSYASTHSSDSEBEEBEEBEEAAPPQGGQMSLLGPABKRLPLHSTYTKDGGPGGKAP 2760  
 QY 2790 WPDGFTTAKSSGNGAPBERLRENGDALSRBSGLPPLGSSAOPHKGILLKCCUPTTISE 2849  
 DB 2761 WPDGFTTAKSSGNGAPBERLRENGDALSRBSGLPPLGSSAOPHKGILLKCCUPTTISE 2820  
 QY 2850 KSLILARLPLEOCTGSSRGSASBGRGPPPPPPROSIQOBLNGMPLAMSTIKAGTUNE 2909  
 DB 2821 KSLILARLPLEOCTGSSRGSASBGRGPPPPPPROSIQOBLNGMPLAMSTIKAGTUNE 2880  
 QY 2910 DSSGSEFLFFNFIH 2923

DB 2881 DSSGSEFLFNFLH 2894

RESULT 6  
AAU74826  
ID AAU74826 standard; protein; 2936 AA.  
AC AAU74826;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Human REPT9 9 protein.

REPT9; human; antiinflammatory; cyostatic; immunosuppressive;  
antiviral; anti-HIV; antiaerthetic; anticoagulant; nootropic;  
neuroprotective; antiallergic; antibody; immunogen; endometriosis;  
gastrintestinal disorder; gastritis; oesophageal carcinoma;  
Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
endocrine disorder; hypothyroidism; Kallman's disease;  
autoimmune disease; inflammatory disease; infertility; receptor;  
acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
allergy; osteoarthritis; diabetes mellitus; multiple sclerosis;  
systemic lupus erythematosus; cell proliferative disorder;  
cancer; developmental disorder; Duchenne muscular dystrophy;  
Becker muscular dystrophy; neurological disorder; epilepsy;  
Alzheimer's disease; Huntington's disease; reproductive disorder.

Homo sapiens.  
XX  
PN WO200198354-A2.  
PD  
XD 27-DEC-2001.  
XX  
PF 21-JUN-2001; 2001WO-US19942.  
XX  
PR 21-JUN-2000; 2000US-214027P.  
PR 25-AUG-2000; 2000US-228045P.  
PR 12-DEC-2000; 2000US-255104P.  
XX  
PA (INCYTE GENOMICS INC.  
PI Griffin JA, Kallick DA, Tribouley CM, Yue H, Nguyen DB, Tang YT,  
PI Lal P, Policky JL, Azimzai Y, Lu DM, Graul R, Yao MG, Burford N,  
PI Hatalla AA, Baughm MR, Bandman O, Patterson C, Yang J, Xu Y,  
PI Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Duggan BM, Lu Y,  
DR MPI: 2002-090432/12.  
DR N-PSDB; ABK15177.  
XX  
PT Twelve human receptors (referred to as REPT9-1 to REPT9-12), useful in  
PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
PT proliferative (e.g. cancer) disorders -  
XX  
PS Claim 53; Page 131-138; 157DP; English.

This invention relates to twelve human receptors cDNA sequences referred to as REPT9-1 to REPT9-12), and the proteins encoded thereby. The proteins of the invention may have anti-inflammatory, cyostatic, immunosuppressive, antiviral, anti-HIV, antiaerthetic, muscular active general, anticoagulant, nootropic, neuroprotective, antiallergic activities. The sequences of the invention may be used to produce REPT9 antisense or antagonists, and the protein sequences may be used to raise anti-REPT9 antibodies. These molecules and the REPT9 polynucleotides and polypeptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma, Crohn's disease, irritable bowel syndrome, ulcerative colitis), endocrine (e.g. hypothyroidism disorder, Kallman's disease), autoimmune/inflammatory (e.g. acquired immune deficiency syndrome (AIDS), rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus, multiple sclerosis, systemic lupus erythematosus), cell proliferative (e.g. cancer), developmental (e.g. Duchenne and Becker muscular dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,

CC	Huntington's disease) and reproductive (e.g. infertility, endometriosis)
CC	disorders. Numerous other examples of each disorder are given in the
CC	specification. The present sequence represents the human RBPt9 protein
CC	sequence of the invention.
XX	
SQ	Sequence    2936 AA;
	Query Match                 98.3%;    Score 15279;    DB 23;    Length 2936;
	Beet Local Similarity      98.4%;    Pred. No. 0;
	Matches 2893;    Conservative    2;    Mismatches    14;    Indels    32;    Gaps    8;
QY	1 MRSPATGVPLEPT-PPPLLILLLLILPPLLLGGDVGFCSRSLSGRGSAGACAPMGWICP    59
DB	1 MRSRTGVPLELPPTPPPLLILLLLILPPLLLGGDVGFCSRSLSGRGSAGACAPMGWICP    60
QY	60 SSASNLWTYSRCDCDAGTELTGHVPHHDGLRWCPSEBAHIPLPPAPEGCPMSCLLGI    119
DB	61 SSASNLMWYSRCDCDAGTELTGHVPHHDGLRWCPSEBAHIPPPAPEGPMSCLRLGI    120
QY	120 GGHLSPQCKTLPRHEHCLKAPRLRCOSCYLAOPGRAGERSPRESISGGRKKNVTAP    179
DB	121 GGHLSPQCKTLPRHEHCLKAPRLRCOSCYLAOPGRAGERSPRESISGGRKKNVTAP    180
QY	180 QFQPSYQAVPENQAPGTVPASLRALDPBEGEAGRLEYTMDALFDRSNQFSLDPYTG    239
DB	181 QFQPSYQAVPENQAPGTVPASLRALDPBEGEAGRLEYTMDAFDERSNQFSLDPYTG    240
QY	240 AVTTAEELDEPKSTHYFRVTAODHGMPRSALATLITLTDTMDHPVPVBQCCKESLR    299
DB	241 AVTTAEELDKSTSHFRVTAODHGMPRSALATLITLTVDNDHPVBEQDSKYKESLR    300
QY	300 ENLEVGVEVLTVRAITDDADAPPNAILLYRLLEGSGSGSPSEVFIEDPRSGLVRTGPPVREE    359
DB	301 ENLEVGVEVLTVRAITDDADAPPNAILLYRLLEGSGSGSPSEVFIEDPRSGVIRTRGPPVREE    360
QY	360 VESYQLTVKASDQGRDPGPSSTTAAVFLSYEDNDNAPOFSKKRYTVQVRBDVTPGAPVL    419
DB	361 VESYQLTVKASDQGRDPGPSSTTAAVFLSYEDNDNAPQFSKKRYTVQVEDVTPGAPVL    420
QY	420 RVTSDDDKSNAVHVSIMSGNARGGYIDAQTGALDVVSPLDYEETTKETYLKVARDG    479
DB	421 RVTSDDDKSNAVHVSIMSGNARGGYIDAQTGALDVVSPLDYEETTKETYLKVARDG    480
QY	480 GRPLSVNSGIVTVQVLINDNAPIFVSTFPQATVLESVPLGYLVHVOAIDADAGDNAR    539
DB	481 GRPLSVNSGIVTVQVLINDNAPIFVSTFPQATVLESVPLGYLVHVOAIDADAGDNAR    540
QY	540 LEYLAQVGDHPPTINNGGWISVAEELREERDPSPCVBARDHGTPALITASASYVT    599
DB	541 LEYLAQVGDHPPTINNGGWISVAEELREERDPSPCVBARDHGTPALITASASYVT    600
QY	600 VLVDVNDNNPTFTQPEYTVRLNEBAAVGTSVVTVSAVDRDAHASYIYQITSGNTRNRPISIT    659
DB	601 VLVDVNDNNPTFTQPEYTVRLNEBAAVGTSVVTVSAVDRDAHASYIYQITSGNTRNRPISIT    660
QY	660 SÖSGGGVLSIALPLDYTLERQYVLAATAASGTQDPAQIVNVNTDANTHAPVQSSHYTV    719
DB	661 SÖSGGGVLSIALPLDYTLERQYVLAATAASGTQDPAQIVNVNTDANTHAPVQSSHYTV    720
QY	720 NVNEDRAGCTTVVLIASITBEDTENAKITYFMEDSIPOFRIADPYGAVTTQAELDIYEDOV    779
DB	721 NVNEDRAGCTTVVLIASITBEDTENAKITYFMEDSIPOFRIADPYGAVTTQAELDIYEDOV    780
QY	780 SYTLATIAARDNGIPÖKSDDTYTLBLVNDVANDNAPÖPLRDYSÖGSVYEDVPPFTSVIOISA    839
DB	781 SYTLATIAARBNGIPÖKSDDTYTLBLVNDVANDNAPÖPLRDYSÖGSVYEDVPPFTSVIOISA    840
QY	840 TDRSGANGVAVFTFOCGDDGDDGFIVESYSGIVRTLRLDRENVAQYVRAAYADKMP    899
DB	841 TDRSGANGVAVFTFOCGDDGDDGFIVESYSGIVRTLRLDRENVAQYVRAAYADKMP    900
QY	900 PARTMEVTVVLVDVNDNPVFQODEDFVBEENSPIGLAVARVATADPDGNTAOIMYO    959

Db 901 PAATMEVTVTVLVDVNDNPEVFEQDEDFVEENSPGLAVARVATADPDDEGTNAQIMYQ 960  
 QY IVEGNIPEVFOLDIFSGELTALVDLDEDEPERVYLYIQTASAPLVSRAVTHVRLDRND 1019  
 Db 961 IVEGNIPEVFOLDIFSGELTALVDLDEDEPERVYLYIQTASAPLVSRAVTHVRLDRND 1020  
 QY 1020 PVLVGNFELFNNVYTNRSSSPFGAIGRVPAHDPDISLTYSPERGNELSVLNNAST 1079  
 Db 1021 PVLVGNFELFNNVYTNRSSSPFGAIGRVPAHDPDISLTYSPERGNELSVLNNAST 1080  
 QY 1080 GEKLSRALDNNRPLAIVSVSDGVHSTYACALRVITITBMTHTSTTLLEMSRE 1139  
 Db 1081 GEKLSRALDNNRPLAIVSVSDGVHSTYACALRVITITBMTHTSTTLLEMSRE 1140  
 QY 1140 RFLSPILGLFIQAVATLAPPDHVVVFNQRTDAPGHIILNVLSVSGPPGPGGPPF 1199  
 Db 1141 RFLSPILGLFIQAVATLAPPDHVVVFNQRTDAPGHIILNVLSVSGPPGPGGPPF 1200  
 QY 1200 LPSEEDLOERLYNRSLLTAISAQVLPFDNICTLREPCENYMCVSLRFDSSAPFIASS 1259  
 Db 1201 LPSEEDLOERLYNRSLLTAISAQVLPFDNICTLREPCENYMCVSLRFDSSAPFIASS 1260  
 QY 1260 SVLFRIPIHPVGLRCPCBPFTDGCETEVDLCSYRBCGPHRCRREGYTCLCRDGYT 1319  
 Db 1261 SVLFRIPIHPVGLRCPCBPFTDGCETEVDLCSYRBCGPHRCRREGYTCLCRDGYT 1320  
 QY 1320 GEHCEVSASGRCTPGVCKNGGTVMNLYVGFCKDCPSGDFEKPQCQVTRSPASFTT 1379  
 Db 1321 GEHCEVSASGRCTPGVCKNGGTVMNLYVGFCKDCPSGDFEKPQCQVTRSPASFTT 1380  
 QY 1380 FRLGRPFHTLALSPATKERDGLLYNGRFNEKHPFVALEVIQEOVLTFSGESTTV 1439  
 Db 1381 FRLGRPFHTLALSPATKERDGLLYNGRFNEKHPFVALEVIQEOVLTFSGESTTV 1440  
 QY 1440 SPFVPGSVSDGMHTVQLKYNNPILGOTGLPGSBEOKVAVTVVNGCDTGVALRGSVL 1499  
 Db 1441 SPFVPGSVSDGMHTVQLKYNNPILGOTGLPGSBEOKVAVTVVNGCDTGVALRGSVL 1500  
 QY 1500 GNYSCAAOGTQGSKKSLDLTGPIILAGVPLDESPFVRMQFVGCMRLQVDSRIHDA 1559  
 Db 1501 GNYSCAAOGTQGSKKSLDLTGPIILAGVPLDESPFVRMQFVGCMRLQVDSRIHDA 1560  
 QY 1560 DFLANNGTGVCRAKXNVCDNNTCHNGGTVMNQMDAFSCCEPLGFGKSCAQOMANPOIF 1619  
 Db 1561 DFLANNGTGVCRAKXNVCDNNTCHNGGTVMNQMDAFSCCEPLGFGKSCAQOMANPOIF 1620  
 QY 1620 LGSLSVAMHGLSLPIISOPTWLSLMPRTROADGVTLAITGRSTITLQLRBGVMLSVSG 1679  
 Db 1621 LGSLSVAMHGLSLPIISOPTWLSLMPRTROADGVTLAITGRSTITLQLRBGVMLSVSG 1680  
 QY 1680 TGLQASSLRLPERANDGDWHHAQLAGASGPGHAILSPDYGOQARBNLGPRLHGLH 1739  
 Db 1681 TGLQASSLRLPERANDGDWHHAQLAGASGPGHAILSPDYGOQARBNLGPRLHGLH 1740  
 QY 1740 SNTTGGIIPBPAGVARGFRGCLQGVAVSDTPBGVNSLDPSHESINVEGCSLPPCCS 1799  
 Db 1741 SNTTGGIIPBPAGVARGFRGCLQGVAVSDTPBGVNSLDPSHESINVEGCSLPPCCS 1800  
 QY 1800 NPCEPANSYSCNDMDYSVSCSDPGYGDNCTNVCDLNPCEHOSVCTKPSAPHGYTCECP 1859  
 Db 1801 NPCEPANSYSCNDMDYSVSCSDPGYGDNCTNVCDLNPCEHOSVCTKPSAPHGYTCECP 1860  
 QY 1860 NYLAGPYCETRIIDPCPRGMWGHPTGBCNCDVSKGPDPCNKTSGCHCKENHYRPPGSP 1919  
 Db 1861 NYLAGPYCETRIIDPCPRGMWGHPTGBCNCDVSKGPDPCNKTSGCHCKENHYRPPGSP 1920  
 QY 1920 TGLLDCDCTPGSLSRVCDDEBDGOCCKEYIGROCDNRPAJMTNGE-----VNYDS 1975  
 Db 1921 TGLLDCDCTPGSLSRVCDDEBDGOCCKEYIGROCDNRPAJMTNGE-----VNYDS 1978  
 QY 1976 CPRAIEGIMWPRTR-----FGLPAAAPCPKXSF-----GTAVRCHDEHGMPLPMLFNC 2025  
 Db 1979 CPRAIEGIMWPRTR-----FGLPAAAPCPKXSF-----GTAVRCHDEHGMPLPMLFNC 2034

QY 2026 TSITSEELKFAERLQORNESGLDSGRSOOLALLLNATQHTAGYFGSPVKAYOLATRL 2085  
 Db 2035 TSITSEELKFAERLQORNESGLDSGRSOOLALLLNATQHTAGYFGSPVKAYOLATRL 2094  
 QY 2086 AHSTORGGLSATQDVHFTENTLKVSGALLDTANKRMBELIQQTEGTAMLLQHYEVA 2145  
 Db 2095 AHSTORGGLSATQDVHFTENTLKVSGALLDTANKRMBELIQQTEGTAMLLQHYEVA 2154  
 QY 2146 SALAQMHTYISPTITVPNTIVISVRLDKGNFAGALPRYELARGOPPLEFTVLLP 2205  
 Db 2155 SALAQMHTYISPTITVPNTIVISVRLDKGNFAGALPRYELARGOPPLEFTVLLP 2214  
 QY 2206 ESVFRETPEVVRPAPGAEQOEPEELARQRHPELSQGEAAVSIYRTLAGLPHNTDP 2265  
 Db 2215 ESVFRETPEVVRPAPGAEQOEPEELARQRHPELSQGEAAVSIYRTLAGLPHNTDP 2274  
 QY 2266 DKRSILVRPRPIINTPVVISVHDEBELPALDKPVTVOFRLLETERTYPCVFMNS 2325  
 Db 2275 DKRSILVRPRPIINTPVVISVHDEBELPALDKPVTVOFRLLETERTYPCVFMNS 2334  
 QY 2326 ILVSGTGMSARGCEVVRNESHVSCQCNHTSPAVLMDVGRRE-----NGEIL 2374  
 Db 2335 ILVSGTGMSARGCEVVRNESHVSCQCNHTSPAVLMDVGRRE-----NGEIL 2394  
 QY 2375 PLKTLTYVALGVTLAALLTFEFLTLRLRSNQHIGIRNLTAAQLGLVFLGINQAD 2434  
 Db 2395 PLKTLTYVALGVTLAALLTFEFLTLRLRSNQHIGIRNLTAAQLGLVFLGINQAD 2454  
 QY 2435 LPFACTVAILLHFLYLTCSFALILBALHYRALTEVDVNTGEMRFYMLGMVPAFIT 2494  
 Db 2445 LPFACTVAILLHFLYLTCSFALILBALHYRALTEVDVNTGEMRFYMLGMVPAFIT 2514  
 QY 2495 GLAVGLDEPGYGNPFCMLSTYDTLVNSPAGVAVASVFLYTLAARASCAORQGE 2554  
 Db 2515 GLAVGLDEPGYGNPFCMLSTYDTLVNSPAGVAVASVFLYTLAARASCAORQGE 2574  
 QY 2555 KKGPSVGLQPSFAVLLLSATWLLALSNSDITLLFHTLFATCNCIOGPFIPLSYVLSK 2614  
 Db 2575 KKGPSVGLQPSFAVLLLSATWLLALSNSDITLLFHTLFATCNCIOGPFIPLSYVLSK 2634  
 QY 2615 EYRKALKIACSKRPSDDALTTKSTLTSSYNCPSRYADGRLYQPGDASGLHSTSRGK 2674  
 Db 2635 EYRKALKIACSKRPSDDALTTKSTLTSSYNCPSRYADGRLYQPGDASGLHSTSRGK 2694  
 QY 2675 SOPSYIFLAREBSALNPGQGPGLGDPGSLFLBGOODOHPDUTSDSDLSEDDQSGSY 2734  
 Db 2695 SOPSYIFLAREBSALNPGQGPGLGDPGSLFLBGOODOHPDUTSDSDLSEDDQSGSY 2754  
 QY 2735 ASTHSDSEEEEBEERBAAPRGGQWDSLGPAGRLPLHSTPDGPGPGKAPWPGDF 2794  
 Db 2755 ASTHSDSEEEEBEERBAAPRGGQWDSLGPAGRLPLHSTPDGPGPGKAPWPGDF 2814  
 QY 2795 GTTAKESGNGAPREERLBNDDALSREBSLGPLPSSSAQPHKXILKXCLPTISKSSLL 2854  
 Db 2815 GTTAKESGNGAPREERLBNDDALSREBSLGPLPSSSAQPHKXILKXCLPTISKSSLL 2874  
 QY 2855 RLPLEQCTGSSRGSASRSGRPPRPPOSLOBOULNGWPIAMSIXAGTVDEDSGS 2914  
 Db 2875 RLPLEQCTGSSRGSASRSGRPPRPPOSLOBOULNGWPIAMSIXAGTVDEDSGS 2934  
 QY 2915 E 2915  
 Db 2935 E 2935  
 RESULT 7  
 ABB11404 standard; peptide; 2560 AA.  
 ABB11404;  
 11-JAN-2002 (first entry)

XX Human FLAMINGO 1 homologue, SEQ ID NO:1774.  
 DE Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoiesis regulation; tissue growth; immunomodulator; actin;  
 KW inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; anti-inflammatory;  
 KW antiaesthetic; antiallergic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antitumor.  
 XX Homo sapiens.  
 OS WO200157188-A2.  
 PN 09-AUG-2001.  
 PD 05-FEB-2001; 2001WO-US03800.  
 PE 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 PR (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Dmanac RT;  
 PI WPI; 2001-457740/49.  
 DR N-PSDB; ABA08648.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX Claim 20; Page 179-181; 1963jp; English.

CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX Sequence 2560 AA;  
 SQ  
 Query Match 86.2%; Score 13399; DB 22; Length 2560;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 2530; Conservative 1; Mismatches 13; Indels 4; Gaps 3;  
 QY 380 STTAAPVLSVDDMDNAPQSEKRYVQVREDYTPGAPVLRVTASDRDKSNAAVHYSIM 439  
 DB 13 STTAAPVLSVDDMDNAPQSEKRYVQVREDYTPGAPVLRVTASDRDKSNAAVHYSIM 72  
 QY 440 SGNARQGYLDAGQALDVSPLDYETTKETTLVRADQGGRPPLSNVSGLVTVQVLDIN 499  
 DB 73 SGNARQGYLDAGQALDVSPLDYETTKETTLVRADQGGRPPLSNVSGLVTVQVLDIN 132  
 QY 500 DNAPIPVSTPQATVLSVPLGYLVHVOAIDADAGNARLEVRLAGVGHDPPTINNGT 559  
 DB 133 DNAPIPVSTPQATVLSVPLGYLVHVOAIDADAGNARLEVRLAGVGHDPPTINNGT 192  
 QY 560 GMSVAAELDRREYDFYSFGVEARDHGTALTSASVTVLVDNNDNPTFOPEYVRL 619  
 DB 193 GMSVAAELDRREYDFYSFGVEARDHGTALTSASVTVLVDNNDNPTFOPEYVRL 252  
 QY 620 NEDAAVGSVTVTSAVDRDAHSVTTYOITSGTNRPSITSQSGGLVSLALPLDYLER 679  
 DB 253 NEDAAVGSVTVTSAVDRDAHSVTTYOITSGTNRPSITSQSGGLVSLALPLDYLER 312  
 QY 680 QYLAATASDGTRODTQIIVNVTDANTHRPVFQSSHVTVNVNDRPAGTTVLISATDE 739  
 DB 313 QYLAATASDGTRODTQIIVNVTDANTHRPVFQSSHVTVNVNDRPAGTTVLISATDE 372  
 QY 740 DTGNARITTFMEDSIPQFRIDADTGAVTTOAELDYEDOVSYTLATARDNGIPQKSDTT 799  
 DB 373 DTGNARITTFMEDSIPQFRIDADTGAVTTOAELDYEDOVSYTLATARDNGIPQKSDTT 432  
 QY 800 YLETLVNDVNDNAPQFRIDSVQSVYEDVPPFSTVLOISATDRSGANGRVFTFOGDD 859  
 DB 433 YLETLVNDVNDNAPQFRIDSVQSVYEDVPPFSTVLOISATDRSGANGRVFTFOGDD 492  
 QY 860 GDGDFVSTSGVIRTLRLDRRENVAGYLRAYAVDGMPPARTPMETVTVLVDNNDP 919  
 DB 493 GDGDFVSTSGVIRTLRLDRRENVAGYLRAYAVDGMPPARTPMETVTVLVDNNDP 552  
 QY 920 VFEQDEDFVBEENSPIGLAVARVATDPDEGNAQIMYOIVBGNIPBEVFOIDIFSGELT 979  
 DB 553 VFEQDEDFVBEENSPIGLAVARVATDPDEGNAQIMYOIVBGNIPBEVFOIDIFSGELT 612  
 QY 980 ALVDLDYEDDEBEVYLVYQASAPLVSRATVTHNLLDRNDNPPVLGNBEILFNNTVTRSS 1039  
 DB 613 ALVDLDYEDDEBEVYLVYQASAPLVSRATVTHNLLDRNDNPPVLGNBEILFNNTVTRSS 672  
 QY 1040 SFPFGAIGRAPADPDSDSLTYSFPERGNEISVTLNASTGBELKLSALDNRPLEAIMS 1099  
 DB 673 SFPFGAIGRAPADPDSDSLTYSFPERGNEISVTLNASTGBELKLSALDNRPLEAIMS 732  
 QY 1100 VLVSDDGVHSTVTAOCALRVITITDEMLTHSITTLRLBDMSPERPLSPILGLFTQAAVATLAT 1159  
 DB 733 VLVSDDGVHSTVTAOCALRVITITDEMLTHSITTLRLBDMSPERPLSPILGLFTQAAVATLAT 792  
 QY 1160 PPDHVVVFNQRPDAPAGHIIIVNSISVQPPRGGPPPLPBBDOERLYLRSSILTAI 1219  
 DB 793 PPDHVVVFNQRPDAPAGHIIIVNSISVQPPRGGPPPLPBBDOERLYLRSSILTAI 852  
 QY 1220 SAQGVLPFPDNNICLRBCENYMCVSLARFSDSAPFIASSSVFRPIHPVGGRCRCRCPG 1279  
 DB 853 SAQGVLPFPDNNICLRBCENYMCVSLARFSDSAPFIASSSVFRPIHPVGGRCRCRCPG 912  
 QY 1280 FTGDYCTEYDLCYSRPCGPHGCRSBRGGYTCLCRDGYTGEHCYVSANSGRCTPGVCNK 1339

```

Db 913 FPGDYCEYERDLCYSRPPCGPHGRCSRREGGYTCLCDGTYGHECEVASASGRCTPCVCKN 972
Qy 1340 GGTGCVLLVGVGFKCDSPSGDFEKPVCQVITRSFPASFTTFRGLRFRFHTLALSPATE 1399
Db 973 GGTGCVLLVGVGFKCDSPSGDFEKPVCQVITRSFPASFTTFRGLRFRFHTLALSPATE 1032
Qy 1400 RDGLLLYNGRFPNEKHDFALEVIQEOVOLTFPSAGESTTVSPVPGVSDGQMTYQLKX 1459
Db 1033 RDGLLLYNGRFPNEKHDFALEVIQEOVOLTFPSAGESTTVSPVPGVSDGQMTYQLKX 1092
Qy 1460 YNKPPLGGTGLPGGBSEOKAVVTVVDCDTGVALRTGSVLGNTSCAAGTGGSKSLDL 1519
Db 1093 YNKPPLGGTGLPGGBSEOKAVVTVVDCDTGVALRTGSVLGNTSCAAGTGGSKSLDL 1152
Qy 1520 TGPLLLGGVDPDPESEPVVRARQFVGCGRNLQVDSRHIDMADFLANNGTVPCCPAKKNVCD 1579
Db 1153 TGPLLLGGVDPDPESEPVVRARQFVGCGRNLQVDSRHIDMADFLANNGTVPCCPAKKNVCD 1212
Qy 1580 SNTCHNGGTCAVQWDAFSCCEPLGFGKSCAQEMANPOHFLGSSLYAMHGLSLPLSQPMY 1639
Db 1213 SKTCHNGGTCAVQWDAFSCCEPLGFGKSCAQEMANPOHFLGSSLYAMHGLSLPLSQPMY 1272
Qy 1640 LSLMFTTRQADGYLLQAITRGRSTTTLQIRBEGHVM SVBGTGLQASSLLBPRANDGM 1699
Db 1273 LSLMFTTRQADGYLLQAITRGRSTTTLQIRBEGHVM SVBGTGLQASSLLBPRANDGM 1332
Qy 1700 HHQQLMAGASGGGHAHLLSFYDQGRABEGLGRLGLHLNNTYVGIKGPAAGVAGRFR 1759
Db 1333 HHQQLMAGASGGGHAHLLSFYDQGRABEGLGRLGLHLNNTYVGIKGPAAGVAGRFR 1392
Qy 1760 GCLQGVRSVDPBGVNSLDPSHGESINVBQCSLPDPCDSNPCCPANSYCSNDMSYSCSC 1819
Db 1393 GCLQGVRSVDPBGVNSLDPSHGESINVBQCSLPDPCDSNPCCPANSYCSNDMSYSCSC 1452
Qy 1820 DPGYVNDNCTNVVDLPCHEGQSVCTRRKBPAPHGYTCECPNTLGPYCEHRIIDQPCRGW 1879
Db 1453 DPGYVNDNCTNVVDLPCHEGQSVCTRRKBPAPHGYTCECPNTLGPYCEHRIIDQPCRGW 1512
Qy 1880 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPSGPTCLLDCYPTGSLSRVCDPE 1939
Db 1513 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPSGPTCLLDCYPTGSLSRVCDPE 1572
Qy 1940 DGQCPCKPGVIGRQCRCDNPFAEVTNNGCEVNVYDSCPAIEAGIWWPRTFGLPAAAPC 1999
Db 1573 DGQCPCKPGVIGRQCRCDNPFAEVTNNGCEVNVYDSCPAIEAGIWWPRTFGLPAAAPC 1632
Qy 2000 PKGSPGTAVHCDHNGMLPBNLFCNCTSTTRSELKGPABRLQNNESGLDSGRSQOLALL 2059
Db 1633 PKGSPGTAVHCDHNGMLPBNLFCNCTSTTRSELKGPABRLQNNESGLDSGRSQOLALL 1692
Qy 2060 RNATQTAGYFSGDVAVQALATRLAHBESTRGFGLSATODVHFENLIRVGSALLDPA 2119
Db 1693 RNATQTAGYFSGDVAVQALATRLAHBESTRGFGLSATODVHFENLIRVGSALLDPA 1752
Qy 2120 NKCHMELIQOTBEGTAMLIQHYEAVASALAQNKRHTYLSPTITVTNIVISVRLDKNP 2179
Db 1753 NKCHMELIQOTBEGTAMLIQHYEAVASALAQNKRHTYLSPTITVTNIVISVRLDKNP 1812
Qy 2180 AGAKLPRYALRGEOPDLETTVILPESVRETPPVVAPRGPBEABEPZELARORRHP 2239
Db 1813 AGAKLPRYALRGEOPDLETTVILPESVRETPPVVAPRGPBEABEPZELARORRHP 1872
Qy 2240 LSGGEAVASYIIVRTLAGLLPHNYDPDKSLRVPKPIINTPVVSLSVHDEBELPRALD 2299
Db 1873 LSGGEAVASYIIVRTLAGLLPHNYDPDKSLRVPKPIINTPVVSLSVHDEBELPRALD 1932
Qy 2300 KPVTVQPRLLLETERTPICYFMMHSTLUSGTGMSARGCEVYFRMESHVSQCCNMHTSF 2359
Db 1933 KPVTVQPRLLLETERTPICYFMMHSTLUSGTGMSARGCEVYFRMESHVSQCCNMHTSF 1992
Qy 2360 AYLMADVSRRENGEILLPKTLTYVALGVTLAALLTPPFLTLILNSNOHIGIRRNLTAL 2419

```

```

Db 1993 AYLMADVSRRENGEILLPKTLTYVALGVTLAALLTPPFLTLILILNSNOHIGIRRNLTAL 2052
Qy 2420 GLAOLVFLGINADLPACTVIALILHFLYLCPSNALLBALHLYRLATEVRDNTGPM 2479
Db 2033 GLAOLVFLGINADLPACTVIALILHFLYLCPSNALLBALHLYRLATEVRDNTGPM 2112
Qy 2480 RPYVVLGMVPAFTTGLAVGLDPBGYGNPDFCWLSTYDTLWSPFAPVAVASVFLYI 2539
Db 2113 RPYVVLGMVPAFTTGLAVGLDPBGYGNPDFCWLSTYDTLWSPFAPVAVASVFLYI 2172
Qy 2540 LAARASCAORQGBRKGPVSGLOPSFVYVLLLSATWLLALLSVNSDTLLPHYLPATNC 2599
Db 2173 LAARASCAORQGBRKGPVSGLOPSFVYVLLLSATWLLALLSVNSDTLLPHYLPATNC 2232
Qy 2600 IQGPIFLSYVVLSEKRVKALKACSRKPSDPALTTKSTLTSYNCPSPYADGRLXOPY 2659
Db 2233 IQGPIFLSYVVLSEKRVKALKACSRKPSDPALTTKSTLTSYNCPSPYADGRLXOPY 2292
Qy 2660 GDSAGLSHSTRSGKQPSYIPLLRBSALNPGQPPGLGD-PGSLFLBEG--DDQHP 2716
Db 2293 GDSAGLSHSTRSGKQPSYIPLLRBSALNPGQPPGLGD-PGSLFLBEG--DDQHP 2352
Qy 2717 DT-DSDSLSTEDDQSGSYASTHSDSREBEREEREAFFGEGQMSLLGPAGRLPLH 2775
Db 2353 XTRPDFSLSLEDQSGSYASTHSDSREBEREEREAFFGEGQMSLLGPAGRLPLH 2412
Qy 2776 STPKDGGPGKAPWPGDPTTAKSSGNGAPBERLRNGALSRGSLGPLPGSAPPH 2835
Db 2413 STPKDGGPGKAPWPGDPTTAKSSGNGAPBERLRNGALSRGSLGPLPGSAPPH 2472
Qy 2836 KGLKKKCLPTISEKSLRLPLEQCTSSSRGSSABEGSGPPPPRPSLOBLQNGV 2895
Db 2473 KGLKKKCLPTISEKSLRLPLEQCTSSSRGSSABEGSGPPPPRPSLOBLQNGV 2532
Qy 2896 MPAMSTAGTVDBDSGSEFLPFNPLH 2923
Db 2533 MPAMSTAGTVDBDSGSEFLPFNPLH 2560

RESULT 8
AAB42192
ID AAB42192 standard; Protein, 2405 AA.
XX
AC AAB42192;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1956 polypeptide sequence. SEQ ID NO:3912.
XX
KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
vulnerary; antiparasitic; antiparkinsonian; noctropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypertensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antineumatic; antithyroid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.

```



PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinketsu RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC76401.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 3067-3072; 5507pb; English.

XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteoplastic; anticonvulsant; antiallergic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasoregic;  
 CC antidiabetic; hypotensive; dermatologic; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihypertoid; and antidiabetic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 2405 AA;

Query Match 90.7%; Score 12541.5; DB 21; Length 2405;

Best Local Similarity 97.8%; Pred. No. 0; Mismatches 28; Indels 3; Gaps 2;

Matches 2354; Conservative 23; Mismatches 28; Indels 3; Gaps 2;  
 QY 516 ESVPLGLVLAHVOAIDADAGNARLEVLAVGVHDPFTTINNCGTMSVAELDREYDF 575  
 Db 1 ESVPLGLVLAHVOAIDADAGNARLEVLAVGVHDPFTTINNCGTMSVAELDREYDF 60  
 QY 576 YSFGVEARDHGTPLATASASVTVYLDVNDNPPFTQPEYTVRLMEDAAVGTSTVTSAY 635  
 Db 61 YSFGVEARDHGTPLATASASVTVYLDVNDNPPFTQPEYTVRLMEDAAVGTSTVTSAY 120  
 QY 636 DRDASVTVYQITSGNTNRPSITSSGGGLVSLALPLDYGLAEFOYVLAATASGCTQDT 695  
 Db 121 DRDASVTVYQITSGNTNRPSITSSGGGLVSLALPLDYGLAEFOYVLAATASGCTQDT 180  
 QY 696 AQIVNVTDANTHRPVPFQSSHYTVVNDPRAGTIVVLSATDEDTGENAITYFMEDI 755  
 Db 181 AQIVNVTDANTHRPVPFQSSHYTVVNDPRAGTIVVLSATDEDTGENAITYFMEDI 240  
 QY 756 POFRIADATGATTAELDYEDOVSYTLATARDNGIPQKSDTTYLTLELVNDVNDAPQF 815  
 Db 241 POFRIADATGATTAELDYEDOVSYTLATARDNGIPQKSDTTYLTLELVNDVNDAPQF 300  
 QY 816 LRDSYQSVYEDVPFTSVLQISATDRDGLNGRVFTYFQGGDDGDFIYESTSGIYRT 875  
 Db 301 LRDSYQSVYEDVPFTSVLQISATDRDGLNGRVFTYFQGGDDGDFIYESTSGIYRT 360  
 QY 876 LRLRDRENVAAQYVLAAYAVDKGMPARTPMEVTVTVLDVNDNPPVFEQDEDFVVEENSP 935  
 Db 361 LRLRDRENVAAQYVLAAYAVDKGMPARTPMEVTVTVLDVNDNPPVFEQDEDFVVEENSP 420

QY 936 IGLAARVATATPDDEGTNAQIMQIVEGNIPEVFOIDI FSGELTALVDLDYEDREPEYLV 995  
 Db 421 IGLAARVATATPDDEGTNAQIMQIVEGNIPEVFOIDIFSGELTALVDLDYEDREPEYLV 480  
 QY 996 IQATSAPLVSAATVHVALDRNDNPPVLGNPELLFNNTYNNRSSSPGGAIGRPADPD 1055  
 Db 481 IQATSAPLVSAATVHVALDRNDNPPVLGNPELLFNNTYNNRSSSPGGAIGRPADPD 540  
 QY 1056 ISDSLTYSPERGNELSVLINAATGELKSRALDNNRPLEAIMSVSDGVSATACAL 1115  
 Db 541 ISDSLTYSPERGNELSVLINAATGELKSRALDNNRPLEAIMSVSDGVSATACAL 600  
 QY 1116 RVITITDMLTHSTITLLEDMSPERPLSPILGFIQAAVATATPDPHVVFNQDRTDA 1175  
 Db 601 RVITITDMLTHSTITLLEDMSPERPLSPILGFIQAAVATATPDPHVVFNQDRTDA 660  
 QY 1176 PGCHILNVSLVQPPRGGGPPPLBEDIOERLYINRSULTAISARVLPEDNICLRE 1235  
 Db 661 -SSNINVTESALLPGVGRG--QPPSEDIQEOIYNRTLLTITSTORVLPEDNICLRE 717  
 QY 1236 PCENYMCVSAVLAEPDSSAPPLIASSSVLFRPIHPVGLRCRCPGPTGDYCEVLDLQYR 1295  
 Db 718 PCENYMCVSAVLAEPDSSAPPLIASSSVLFRPIHPVGLRCRCPGPTGDYCEVLDLQYR 777  
 QY 1296 PCGPHGRCSRBEQGYTLCRDGYTHGHCYVANSRGCTPGVCKNGGTCVNLVGGFKDC 1355  
 Db 778 PCGANGCREREGGYTLCRDGYTHGHCYVANSRGCTPGVCKNGGTCVNLVGGFKDC 837  
 QY 1356 PSGBFEKPYQVNTSRPASPTFRGLRORPHFTLALSPATXRDELILYNGFNRKH 1415  
 Db 838 PSGBFEKPYQVNTSRPASPTFRGLRORPHFTLALSPATXRDELILYNGFNRKH 897  
 QY 1416 FVALEVLQEOVOLTFASGESTTVSPVPVGVSDQWHTVOLKTYNRPPLGQTLPOGPS 1475  
 Db 898 FVALEVLQEOVOLTFASGESTTVSPVPVGVSDQWHTVOLKTYNRPPLGQTLPOGPS 957  
 QY 1476 EOKYAVTVVDCDGTVALRFGSVLGNYSCAAQGTGGSKSLDITGPLLIGVVDLPESF 1535  
 Db 958 EOKYAVTVVDCDGTVALRFGSVLGNYSCAAQGTGGSKSLDITGPLLIGVVDLPESF 1017  
 QY 1536 PYRRROFVGCMRLQVDSRHIDMADFIANNGTVPGCGRKKNVCDSNCHNGGTCVNMWDA 1595  
 Db 1018 PYRRROFVGCMRLQVDSRHIDMADFIANNGTVPGCGRKKNVCDSNCHNGGTCVNMWDA 1077  
 QY 1596 FSCCEPLFGGKSCAQMANNPQHLGSSLVAMHGLSLPISQPMWLSIMFRTQADGVLQ 1655  
 Db 1078 FSCCEPLFGGKSCAQMANNPQHLGSSLVAMHGLSLPISQPMWLSIMFRTQADGVLQ 1137  
 QY 1656 AITRGASTITLQLRBGMVLSVSGTGLQASSLRLBPRANDGWMHQAQLAASGGPGHA 1715  
 Db 1138 AITRGASTITLQLRBGMVLSVSGTGLQASSLRLBPRANDGWMHQAQLAASGGPGHA 1197  
 QY 1716 ILSFDYQOARAEQNLGRLHGLHSNTYVGIIGPAGGVARGGCLQGVRSPTPEGVN 1775  
 Db 1198 ILSFDYQOARAEQNLGRLHGLHSNTYVGIIGPAGGVARGGCLQGVRSPTPEGVN 1257  
 QY 1776 SLDPSSHGESINVEQGSLLPDCSNPCPNANSYCSNMDYSYSGCDPGYGDNCTNVCDLN 1835  
 Db 1258 SLDPSSHGESINVEQGSLLPDCSNPCPNANSYCSNMDYSYSGCDPGYGDNCTNVCDLN 1317  
 QY 1836 PCRHQSVCTKRPAPAGHYTCECPNTYIGPYCEFRIDOPCRGWMGHPTGCPNCNDVSKGF 1895  
 Db 1318 PCRHQSVCTKRPAPAGHYTCECPNTYIGPYCEFRIDOPCRGWMGHPTGCPNCNDVSKGF 1377  
 QY 1896 DPDCNKTSGSCCHKENHYRPGSPPTCLLDCYPTGSLSRVCDPEDDGCPKPGVIGQCQ 1955  
 Db 1378 DPDCNKTSGSCCHKENHYRPGSPPTCLLDCYPTGSLSRVCDPEDDGCPKPGVIGQCQ 1437  
 QY 1956 RCDNPFALVTTCGVNYSDBPRAIEAGIWPPTRFGLPAAACCPKGSFGTAARHCDHR 2015  
 Db 1438 RCDNPFALVTTCGVNYSDBPRAIEAGIWPPTRFGLPAAACCPKGSFGTAARHCDHR 1497  
 QY 2016 GMLPNTFNCTSTTSPBLKGFARLQRBESGLDSGSQLALLLRNATQHTAGVFGSDVK 2075

Db 1498 GMLPEVLFCTSTSTFELKGPAPERLQPNSSGLDSGSOQLALLRLNATHTGYPESDVK 1557  
Qy 2076 VAVQATRLRLAHSTORFGLSATQDVHTEMLRLRGASALDPTANRHEMLQOEGGTA 2135  
Db 1558 VAVQATRLRLAHSTORFGLSATQDVHTEMLRLRGASALDPTANRHEMLQOEGGTA 1617  
Qy 2136 WLLQHYEAVASALAQMRHTYLSPTITVTPNIVISVRLDKGNFAGAKLPRYEALGEOP 2195  
Db 1618 WLLQHYEAVASALAQMRHTYLSPTITVTPNIVISVRLDKGNFAGAKLPRYEALGEOP 1677  
Qy 2196 PDETTVIIPESEVRETPPVVRPAGGEAEPEELARRQRHELSQGEAVASVIIYRTL 2255  
Db 1678 PDETTVIIPESEVRETPPVVRPAGGEAEPEELARRQRHELSQGEAVASVIIYRTL 1737  
Qy 2256 AGLLPHNYDPDKASLKVPKPIINTFVVSISVADDEELLPRALDKVTVQFRLTEERT 2315  
Db 1738 AGLLPHNYDPDKASLKVPKPIINTFVVSISVADDEELLPRALDKVTVQFRLTEERT 1797  
Qy 2316 KPICTVFNHSLIVSGTGSARCCVFPNESHVSCQCNHMTSFVAVLMDVSRRENGEILP 2375  
Db 1798 KPICTVFNHSLIVSGTGSARCCVFPNESHVSCQCNHMTSFVAVLMDVSRRENGEILP 1857  
Qy 2376 LKLTETVALGVTLAALLTFEFLTLRLIRSNQHGIRBNLTALGLAOLVFLGIMQADL 2435  
Db 1858 LKLTETVALGVTLAALLTFEFLTLRLIRSNQHGIRBNLTALGLAOLVFLGIMQADL 1917  
Qy 2436 PFACVIAIILHFLYCTFSWALLLEALHLYRALTEYDVNTGPMRYMIMGVPAFITG 2495  
Db 1918 PFACVIAIILHFLYCTFSWALLLEALHLYRALTEYDVNTGPMRYMIMGVPAFITG 1977  
Qy 2496 LAVGLDPEGYGNDFCMLSTYDTLINSFAGVPAFVAVSMGVFYTLTAAASCAAQROGFEK 2555  
Db 1978 LAVGLDPEGYGNDFCMLSTYDTLINSFAGVPAFVAVSMGVFYTLTAAASCAAQROGFEK 2037  
Qy 2556 KGPVSGLOPSFAVLLLSATWMLALSVNSDTLPHYLPATCNCIGCPFLISTVYVLSRE 2615  
Db 2038 KGPVSGLOPSFAVLLLSATWMLALSVNSDTLPHYLPATCNCIGCPFLISTVYVLSRE 2097  
Qy 2616 VRKALFLACSRKSPDPALTTKSTLTSYNCSPYADGRLYQYGDASGLHSTRSGKS 2675  
Db 2098 VRKALFLACSRKSPDPALTTKSTLTSYNCSPYADGRLYQYGDASGLHSTRSGKS 2157  
Qy 2676 QBSYITFLLREBESALNPGQPGGLGDPGSLFLEGQOQOHPDPTDSDSLBDOSGSA 2735  
Db 2158 QBSYITFLLREBESALNPGQPGGLGDPGSLFLEGQOQOHPDPTDSDSLBDOSGSA 2217  
Qy 2736 STHSSE 2795  
Db 2218 STHSSE 2277  
Qy 2796 TTAKESGNGAPBERLRENGDALSRBSGLPLPGSSAOPHKGILLKKCLPTISEKSLLR 2855  
Db 2278 TTAKESGNGAPBERLRENGDALSRBSGLPLPGSSAOPHKGILLKKCLPTISEKSLLR 2337  
Qy 2856 LPLEOCTGSSRGSSABSGSGRPPPPRQSLQEOQLNGMPFAMSITKXTDDESSGE 2915  
Db 2338 LPLEOCTGSSRGSSABSGSGRPPPPRQSLQEOQLNGMPFAMSITKXTDDESSGE 2397  
Qy 2916 FLFFNFLLH 2923  
Db 2398 FLFFNFLLH 2405

RESULT 9  
AAU68533 standard; Protein; 3014 AA.  
ID AAU68533;  
AC AAU68533;  
XX  
XX 16-JAN-2002 (first entry)  
XX  
XX Human novel cytokine encoded by cDNA 790CIP2C\_4 #1.

XX Human; cytokine; cell proliferation; cell differentiation;  
KW antiinflammatory; stem cell growth factor; activin; inhibin; cancer;  
KW nervous system disease; neuropathy; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; spinal cord disorder;  
KW head trauma; stroke; myeloid cell disorder; lymphoid cell disorder;  
KW platelet disorder; thrombocytopaenia; stem cell disorder;  
KW aplastic anaemia; tissue regeneration; wound healing; ulcer;  
KW osteoporosis; osteoarthritis; bone degenerative disorder;  
KW periodontal disease; fibrosis; reperfusion; immune disorder; SCID;  
KW severe combined immunodeficiency; infection; autoimmune disorder;  
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; allergy;  
KW asthma; coagulation disorder; haemophilia; sepsis; nephritis;  
KW inflammatory bowel disease; food supplement; immunogen.  
OS Homo sapiens.  
XX  
XX WO200175093-A1.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US10484.  
XX  
XX 31-MAR-2000; 2000US--0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PR 22-SEP-2000; 2000US-0668680.  
PR 23-OCT-2000; 2000US-0695618.  
PR 30-NOV-2000; 2000US-0728711.  
PR 14-MAR-2001; 2000US-0728711.  
XX  
XX (HISB-) HISBQ INC.  
PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J, Xu C,  
PI Yang Y, Zaho QA, Chen R, Wang D, Goodrich RW, Liu C, Drmanac RT;  
XX  
XX MPI: 2001-626432/72.  
XX N-PBDB; AAS59825.  
XX  
XX New polypeptides and nucleic acids, useful for diagnosis, treatment of  
PT inflammatory, autoimmune, neurological, myeloid or lymphoid cell, bone  
PT degenerative disorders, cancer and promoting wound healing  
XX  
PS Claim 20; Page 249-254; 336pp; English.  
XX  
XX The invention relates to isolated human polypeptides (which may be  
CC cytokines) and the polynucleotides encoding them. The proteins are useful  
CC for identifying a compound which binds to it (e.g. modulators, agonists  
CC and antagonists). The polynucleotides are useful as an array for mismatch  
CC detection. The proteins and nucleic acids are useful as nutritional  
CC sources or supplements. The protein exhibits exhibits activity relating  
CC to cytokine, cell proliferation, cell differentiation, antiinflammatory,  
CC stem cell growth factor activity, immune stimulating or immune  
CC suppressing and activin or inhibin related activities. The proteins (and  
CC antibodies raised against them) and nucleic acids are therefore useful in  
CC the diagnosis and treatment of diseases and disorders such as cancer,  
CC central and peripheral nervous system diseases and neuropathies,  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, spinal cord disorders, head trauma, cerebrovascular  
CC diseases, stroke, myeloid or lymphoid cell disorders, platelet disorders,  
CC thrombocytopaenia, stem cell disorders, aplastic anaemia, for  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, and in tissue repair, healing of burns, incisions, ulcers, for  
CC treating osteoporosis, osteoarthritis, bone degenerative disorders, or  
CC periodontal disease, lung or liver fibrosis, reperfusion injury in  
CC various tissues, various immune deficiencies and disorders including  
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
CC autoimmune disorders (e.g. multiple sclerosis, rheumatoid arthritis,  
CC diabetes mellitus, myasthenia gravis), allergic reactions and conditions,  
CC such as asthma or other respiratory problems, coagulation disorders,  
CC haemophilia, septic shock, sepsis, arthritis, nephritis and inflammatory  
CC bowel disease, viral infection and are useful in altering bodily  
CC characteristics. The present sequence represents a novel protein of the  
CC invention.

XX Sequence 3014 AA;

Query Match 56.3%; Score 8754; DB 22; Length 3014;  
 Best Local Similarity 55.8%; Pred. No. 0;  
 Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;

12 TTPPRLILILILILPRLIGDOYQPCRSIGSRGSSGACAPMGICPSSASNLMLYTSR 71  
 57 TTRAPRELL-----DVGDRGRLAGR--RRVSGAGRP-----PLQVLY----- 93

72 CRDAGELTGHVPHHIDLRWVCPSEBAHPLPRAPBEGCPMSCRLIGGHSPOGKLT 131  
 94 ABAATPALSRL-----RARKLP-----GCAPARLCGTARLC--GALCF 133

132 P-----BEAPCLKA-----PLRCS-----CKLAQAPG- 155  
 134 PVGGCAAAQHSALAAPTLPRACRCPRRPRCPGAPICLPFGSGVRLRLCALARAAGA 193

156 LRAG-----ENSPESLGGRRKRVNTAP---QFQPSYQAT 189  
 194 VAVGALBAATAGTSPSPSPPLPPLPEARACPARARRGTGSGSLKFPMPNYQA 253

190 VENQPACTPVASLRAIDPDEBAGRLBYTMDALPDSRNOFPSLDPVTGATTTABEIDR 249  
 254 LPENEAGTLILQLHAYTIBGBBEKVSYTMGLPDESRGYFRIDSATGAVSTDSVDR 313

250 ETKSTHVFVTAQDHGMPPRSALATLTILVTDTNDHPVFEQOEYKESLREMLBYGYEVL 309  
 314 ETKEHVLRKVAVDSTPRSATYTYTVLVKOTNDHSPVFESEYRERERREMLBYGYEVL 373

310 TVRATDGDAPRVANILYRLLESGSGSPSEVFEIDPDSGYIRTRGPVDRBEVESYQTVEA 369  
 374 TIRADDRSPRIANLRVYLGAM---DVFOLESSGVSTRAVLDRREAEYQLYVEA 429

370 SPOGDRPGRSTTAALVFLVEDNDNAPQFSKRYVVOVREDVTGAPLYRTTASDRXG 429  
 430 NDOGRPGSLATATVYIEVDENDYPOFSQNVVOVREDVGLNTAALRQATDRDG 489

430 SNAVVHYSIMSGNARQFYLDAQTGALDVSPLDVETTKYTLRAVRAODGSRPLSNVSG 489  
 490 QNAALHYSILSGNVAQFYLHSLSGIDVYNLPDPEDVQKSLSKAQCGSRPLINSVG 549

490 LVTVOVLINDNAPFVSTPFOATLBSVPLGLVLAFOALDADAGNARLEYRLA---- 545  
 550 VVSVOVLVDNDEBPIFVSSPFOATVLENVPLGYPVNHIOADVADSGENARLHYRLVDAS 609

546 ---GVGH-----DEPTINNGTGMISVAABLDREEDVFGSGVAPRHGTALTA 592  
 610 TFLGGGAGPKNPAPLPDPFPQIHNSGMITVCAELDREVEHYSIGVAADHSGSPMS 669

593 SASVSVTVLVDNNDNFTQPEYTVLALNEDAAVGSVTVASVADDAHSVITYQITSGNT 652  
 670 STSVSITVLDVNDNBPVFTQPTBELRLNDAVGSVTLQARBDANSVITYQLTGNT 729

653 RNRFSITSGSGGLYSALPLDYKERYVLAATVMSDGTRODTAQIVNVTANTHPRVF 712  
 730 RNRFLSSQGGGLTLPLPLDYKQEQYVLAATVMSDGTSTAHVILVNTANTHPRVF 789

713 QSSHTVANNEDRPAGTIVLLSATDEDEGENARITYFMEDSIPOFRIDADGAVTTQAE 772  
 790 QSSHIVSVSEDRPVGSTIATISANDEDEGENARITYVQDPQFRIDPDGNTYME 849

773 LDYEDQVSTTLAITARANGIPQKSDITYLEILVNDVNDNAPOFLDSYQSYEDPPT 832  
 850 LDYENQVATLTITMADNGIPQKSDITTLILILANDANAPOFLMDPYGSIFFEDAPST 909

833 SYLQISATRDGSLANGRVYTTQGGDDGDDTVESTSGIVTTLRLDRENAQVLYAY 892  
 910 SILQVATRDGSGPNRLLYTFQGGDDGDDPYIBTSGVIRTORRLDRENAVNLNAL 969

893 AYDKGMP--PATTMEVTVVLDVNDNBPVFEODEPVPVEENSPIGLAVARTATDPDEG 951

970 AYDRGSPPLLSAIVEIQTLLINDNAPMEKDEBLEFVEENNPVGSVVAKIRANDPDEG 1029

952 TNAQIMQIVEGNIPVEFQLDIPSGELTALVDLYEDRPEVYLVLOATSAPLVSRATYHV 1011  
 1030 PNAQIMQIVGDMKHFQQLDLANDGLRAMELDEPVREREVLYVQAASALVSAATYHI 1089

1012 RLDRNDNPPVLANFELLFNNTVNRSSPPGGAIGRVAPANDPDISLSTYSEFGENELS 1071  
 1090 LLYVQNDNBPVLPFQQLFNNTVNRKSNSPFTGVYGCIPADHPVDSILANTYFVQGNELR 1149

1072 LVLANASTGELKLSALDNNRPLAISMVTVSDGVSTTAQACALVYITIDEMTLHSTL 1131  
 1150 LLLLDPAIGELQSLHDLNNRPLRALMEVSDGHSHTATCTLRVITITIDMLNSTITV 1209

1132 RLEDMSPERFLSPILGLFIOAVATLATPDPHVAVFVNOQDPTDABGHIILNVSVGQPP 1191  
 1210 RLENNQSEKFLSPILALFVEBVAVLSTKQDVVFVNVQNTDV--SSILNVTSALLPG 1268

1192 GPGGPPFLPSEBDYQERLYNRSLITAI SAQVLPFDNITCLREPCENYMRCSVYLRPDS 1251  
 1269 GVRG--QFPSEBDYQERLYNRSLITAI STQVLPFDNITCLREPCENYMRCSVYLRPDS 1326

1252 SAPFLASSVLPFRPIHPVGRRCRCPGFTGDCETVDLCYSRPGSGHRCRSRGGYT 1311  
 1337 SAPFLSSTVLPFRPIHPINGRCRCPGFTGDCETVDLCYSRPGSGHRCRSRGGYT 1386

1312 CLCRDGYTGEHCEVASASGRCTPGVCKNGGTCVNLVVGAFKDCDPSGDFEKPQCVTTRS 1371  
 1387 CECFEDFTGEHCEVDASGRCAVCGNNGGTCVNLVVGAFKDCDPSGDFEKPQCVTTRS 1446

1372 FPAHSFTFRGLRQRFHTLALSPATYERDGLILYNGRFNEKHPVAVLEVIQEOVOLTES 1431  
 1447 FPPQSFVTFRGLRQRFHTLALSPATYERDGLILYNGRFNEKHPVAVLEVIQEOVOLTES 1506

1432 AGESITTVSPVPGVSDGCHTVOYKYNKELLGQGLPGSBEQKAVAVTVVQGCDDTV 1431  
 1507 AGESITTVAPRVPGVSDGCHTVOYKYNKELLGQGLPGSBEQKAVAVTVVQGCDDTV 1566

1492 ALRPGSVLGNTSCAAGTGGSKSLDTGPTLLGVPDLPESPPVNRQFVGCGRNLQV 1551  
 1567 AVRFQKIDIGNSCAAQGTQGSKSLDTGPTLLGVPDLPESPPVNRQFVGCGRNLQV 1626

1552 DSRHIDNADPIANNGTVPGPCAKKQVCDSTCHNGGTCVNOVMDAFSGCEPLGFGSKCAQ 1611  
 1627 DGKVDNAGFTANNGTGEGCAARNFCGRRCQNGGTCVNMNMNYLCECPRLFQKCEQ 1686

1612 EMANPOHLLGSLVAMHGLSLPISOPWTLSPMTROADGVLLAITRGSTITLQLEBG 1671  
 1687 AMPHQLPFGSIVSWSDNLIIISVPTLGLMFRIRKEDSVLMBATYSGGPTSRILQILAN 1746

1672 HVMLSVEGTGLQASLSRLRPGRNDGDMHHAQLAL---GASGGGPHAT--LSFDYGOQRAE 1727  
 1747 YLQFVESHGCPDVBSVMLSGRLVTDGEMHLLILKXVKEDSEMKHLVMTMLTGMDNKK 1806

1728 GNLCERLHGLHLSNITVGGIIPGAGVARGFRGLQGYRVSDTPEGVNSLDPHSGESLNV 1787  
 1807 ADIGKMLPGLTVRSVVVVGASLEDKYSVARGFRGCMQGRMGGTPTNVAATLMMNNALKRV 1866

1788 EOGCSLPDPCDSNCPANSYCSNMDYSGSCDGYGDNCTNYCDLNPCHSGHSCTRKP 1847  
 1867 KDCCDVDDPCTSSCPFPNSRCHDAMEDYSVCVCDKGYLAINCVDAKHLNPGCNMGACVSP 1926

1848 SAPHGTYCECPNYLAGPYCETRIDOPCRGMWGHPTGCPNCNDVSKGFPDCKNTSGECH 1907  
 1927 GSPQGYVCEGSPSHYGPYCEKMDLPYCRGMWGNPVGPGCHCAVSKGFPDCKNTNGCQ 1966

1908 CKENHYRPGSPPTCLDCYTGSLSRVCPBEDQCCPKRVYTRQCDRCNNPAPAYTTN 1967  
 1987 CKENYTKLAODTCLPDCDFHSGHSRTCDMATQCAKCPGVIRQCNRCNNPAPAYTTL 2046

1968 GCEVYVDSCPRAISAGIIMPPTRFGLPAAAPCPGSGSGTAVRHGDBERGMPLPMLFNCTS 2027  
 2047 GCEVYVDSCPRAISAGIIMPPTRFGLPAAAPCPGSGSGTAVRHGDBERGMPLPMLFNCTS 2106

```

QY 2028 ITFSEIKGAEERLOENESGLDSRGOQLALLLNATQHTAGYFGSDYKVAVOLATLRLH 2087
DB 2107 ISPVDRANNEKLSRNETQVDGARALQVRLALRSALQHGTCLFGNVRARAYQLGVLQH 2166
QY 2088 ESTQRFGLSATQDVHFTENLELVSGALLDTANKRMWELIQTEGGTAMVLQHTYVASA 2147
DB 2167 ESQOQFDLAATQDAPFHEDEVINHSGLALLAPATRAWEOIQSEGGTAOLRLRLECYFN 2226
QY 2148 LQNMHTLSPPTIYTPNIVISVRLDGNPAGALPRVYALRGSPDLETTVLPS 2207
DB 2227 VANNVRRITLRFPVITANNITLAVDIPDFENFTGAAPVPRDTHSEFPRLSSVSFPD 2286
QY 2208 VFR---ETPVVRPAG-----PGEAQBPEELARORRHPRLSOGAEAVAVIYRT 2254
DB 2287 PFRPPEKGPILRPAGRRTPQTRPGQTEREARISRRRRPDQAQPAVALVIYRT 2346
QY 2255 LAGLLEPHNTDPDKRSIRVPRPIINPVVISVHDEBELPRALDKPVTVQRLTBER 2314
DB 2347 LQQLLEPERYDPPRRSIRLPRPIINPMVSTLVYSGAELPRLEPRVLEFALLEVER 2406
QY 2315 TKPICVFMHSLIVSGTSGMARGCEVPRNESHVSCCNHMTSPFVLMDSVRRENGEL 2374
DB 2407 TKPVCFMHSILAVGGIGMSARGCELSSNRTHVACQCSHTASPAVLMDSIRRENGEVL 2466
QY 2375 PLKTLTYVALGVTALALLTFPFLTLRLRSNQHGRIRNLTALGLAQVLLGLINQAD 2434
DB 2467 PKIIVTYAASVSLALVAVLVLSTVRLMSLHSHHGLAVLPSQLVIVIGINQRE 2526
QY 2435 LPPACTIVAILHLFVLTCTPSMALLEALMLYRALTVRDVNTGPMRFTYMLGMPAPIT 2494
DB 2527 NPFLCTVVAIILHLYIMSTFAMTLVSVLTVYVRLTEVRNIDGPMRFYVVGIGIPALIT 2586
QY 2495 GLAVGLDPGSGYGNPCWLTSTVYTLWSPFAGPAPAVMSVFLYLAARSCAQRQGRE 2554
DB 2587 GLAVGLDPGSGYGNPCWLTSTVYTLWSPFAGPAPAVMSVFLYLAARSCAQRQGRE 2646
QY 2555 KKGVPVSGLOPSPAVLLLSATWMLALLSVNSDTLLFHLFATCNCIQGFPIFLSYVLSK 2614
DB 2647 KKGIVSLKTAFLILLISATWMLGLAVNRDLSPHYLPALPSGLQGFVILLPHCVLQ 2706
QY 2615 EYRKALK-LACSKKEPPDALTKST-LTSYVCEPPIYADG--RLYQPYGDAAGLSHRS 2670
DB 2707 EYRKALKGVLAGKRLLEDSATTATLRLRSINCNTTFDGGPDMLTDLGSESTASLDSIV 2766
QY 2671 RS-----GKSQPSYIPFLAREBSALNPGQPPGKDGPSLFLREGQDQ 2713
DB 2767 RDBGLOKLGVSGLVNGSHGRPASLMPNSCD-----PRG----- 2802
QY 2714 HDPDTDSDDLSTEDQSGSYASTHSSDEEBEERBEAAPPREGWMDSLGPGAERLP 2773
DB 2803 --HDSQSDSELST-DRQSSSYASHSSDSRDDGVGAEEK-----MPARG-----A 2845
QY 2774 LHSTPR-DGPGRGKAPWPDDECTYAKESGNAP-----EERLRENGDL 2818
DB 2846 VHSPTPGADVANNVPRGWPQDQSLAESDSBDPSKPLKYETKVSVEALREBQSHRGKRP 2905
QY 2819 SREGSGAPLPGSSAQP---HKGILKKKCL---PTISEKS---SLTLPLEQCTGSSRSGS 2869
DB 2906 PDESSEGAARLASQPPREGKGLKKKVVYPPVLTITBQTLKRLABKALDEQSFSTSR 2965
QY 2870 ASBGRSGP---PPRPPRQSLQEQLNQVMPJAMSIKAGTVDESSGSE 2915
DB 2966 TSSLSGSGPDCATITVKSFGREPRDHLNGV---ANNVRTGSAQADGSDSE 3012

```

RESULT 10  
AAU02196 standard; Protein; 3014 AA.

AC AAU02196;  
XX  
DT 26-SEP-2001 (first entry)

```

XX DE Seven-pass transmembrane receptor-like protein, MEM1.
DB XX
KW Seven-pass transmembrane receptor; MEM1; therapeutic; diagnostic; MEM2;
KW human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;
KW female reproductive health; lung disorder; brain disorder; schizophrenia;
KW heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;
KW cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;
KW vision-related disorder; neoplastic pathology; MEM7; MEM8.
XX KM
OS Homo sapiens.
XX PN
DB WO200144473-A2.
XX PD
DB 21-JUN-2001.
XX PP
DB 14-DEC-2000; 2000WO-US33909.
XX PR
DB 14-DEC-1999; 99US-0170564.
XX PR
DB 27-DEC-1999; 99US-0173165.
XX PR
DB 27-DEC-1999; 99US-0173362.
XX PR
DB 04-JAN-2000; 2000US-9965564.
XX PR
DB 09-AUG-2000; 2000US-0223929.
XX PR
DB 13-DEC-2000; 2000US-9965565.
XX PA
DB (CURA-) CURAGEN CORP.
PI Spaderma SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;
XX WPI; 2001-398154/42.
XX DR
DB N-PEDB; AAS06332.
XX PT
DB Novel polypeptide comprising members of protein families (e.g.,
PT seven-pass transmembrane receptor proteins) according to presence of
PT domain and sequence relatedness are useful for treating or preventing,
XX e.g., Alzheimer's and Parkinson's -
XX PS
DB Claim 1, Fig 2; 162pp; English.
XX CC
DB The sequence represents the amino acid sequence of seven-pass
XX transmembrane receptor-like protein, MEM1, selected from a group (MEM1-
XX MEM6) comprising members of protein families according to the presence of
XX domain and sequence relatedness, e.g., seven-pass transmembrane receptor
XX protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein
XX (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-
XX MEM8). The MEM polypeptides (I), nucleic acids (II), and antibodies (III)
XX are all useful for treating or preventing a pathology associated with (I)
XX comprising administering (I), (II), or (III) to a subject (preferably a
XX human). In addition, (I), (II), and (III) may be used to manufacture a
XX medicament for treating a syndrome associated with a human disease that
XX is associated with (I). Furthermore, (I) may be used to identify agents
XX that bind to it, screen modulators of its activity and determine the
XX presence or predisposition to a disease associated with altered levels of
XX (I). Disorders for MEM1 include Alzheimer's or Parkinson's disease,
XX cancer, nephrology, and female reproductive health. Disorders for MEM4
XX include those involving the lung and/or brain (e.g., schizophrenia). For
XX MEM5, disorders include heart (arrhythmic disorders) and other muscular
XX disorders, clotting deficiencies and cobalamin deficiencies (e.g.,
XX pernicious anaemia). Such disorders for MEM6 include diabetes, whereas
XX disorders for MEM7 and MEM8 include vision-related disorders, cancer,
XX and other neoplastic pathologies.
XX SQ
DB Sequence 3014 AA;

```

Query Match 56.3%; Score 8754; DB 22; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  
12 TPEPPLTLLTLLTLPPLIGDQVGPCKRSIGSRGSGSGACAPMGWLCPSASNTMLTYSTR 71  
DB 57 TPAPRRLT-----DVGRRDRLAGR-HRVSGAGRP-----PLQVRLV----- 93

QY 72 CRDAGTELQHLVPHDGLRVCMPESEBARIPLPAPBEGCPCWCRLLIGIGHLSPOGKLTTL 131  
DB 94 ANSAPALSRRL-----RARTHLDP-----GCABARLGGTARLCL--GALCP 133  
QY 132 P-----BEHPCLKA-----PRLRCS-----CKLAAPG- 155  
DB 134 PVPBGCAAHQSHALAPPTILPACRCPPRRPRCPGPPICLPBGGSVRLRLCLARBAQA 193  
QY 156 LRAG-----ERSPEESLGRKKRNVTAB-----QOPPEPYQAT 189  
DB 194 VRVGLALBAAATAGTPASPPSPPLPPNLPBARAGPARARAGTSGRSGSLKPFMPHYQVA 253  
QY 190 VPPNOCAGPVPASLRAIDPDEGAGLEBTMDLFPBSKNQPFSLDPVTGAVTTAEELDR 249  
DB 254 LPENEPAGTLLIQLAHNTTIEGEBEERSYMEGLFPERBSRGYRIRISACGAVSTDSVLDR 313  
QY 250 ETKSTVFVTAADHGMPPRSALATLTLVTDNTDNDPVPEOEOEYKESLREMLEVEYEL 309  
DB 314 ETKETHTLAVKADVSTPPRSATTYITVLVKDINDHSVPVBOSETHREKREMLEVGYEVL 373  
QY 310 TVRADTDGAPPNANILYRLBSGSGSPSEVFEIDPPSGVIRTRGPVDRREVESYQULTVEA 369  
DB 374 TIRASDRDSPINAMLRVYLGAM-----DVPQLBESSGVVSTRAYLADREBAARYQLVYA 429  
QY 370 SPOGDRGRSTTAANVFLSVEDDNDNAPOPSEKRYVQVREDVTPGAPVLTATYASDRKG 429  
DB 430 NOGRRPGLSATATYVIEVEDENDNVPQPSSEONVYVQVPEBDVGLNTAVLTATVOATDRDQ 489  
QY 430 SNAVVYSSIMSGNARGOFLDAQTAGLADVSPLDVETTEKTYTLRYAODGGRPPLSNVSG 489  
DB 490 QNBAIHYSILSGNVAQFYLHLSGLLDYINPLDEVDOKYSLSIAQOGGPPPLINSBG 549  
QY 490 LVTVOVLINDNAPIFVSTPFOATVLESVPLGYLVAQVADADGNARLELYLA---- 545  
DB 550 VVSVOVLVDNDEBEPFVSSPFOATVLENVPLGVPVHTIQAVDADSEGNARLHYRLVDIAS 609  
QY 546 ---GVGH-----DPPFTINNGTGHISVAELDREEDVPTSPGVARHGHPTALTA 592  
DB 610 TFLGSGSAGPKNPAPTPDPFOIHNSSGMITVCAELDREBEVHSFGVAZVHGHGPPMS 669  
QY 593 SASVSVTVLVANDNNPTFQPEYTVRLNEDAAVGTSVVSAVDDASHVITYOITSGNT 652  
DB 670 STSVSEITVLVDNDNPVFTQPTTELRLNDDAAVGSSVLTLOARDDANSVITYOITGANT 729  
QY 653 RNRFSITSGSGGLVSLALPLDYKLERQYVLAVTASDGTRODTAOIIVANNVTANTHRPVF 712  
DB 720 RNRFPALSQRGGGLITLALPLDYKQEQYVLAVTASDGTSHHTAVLINVNTANHRPVF 789  
QY 713 OSSHTVANNEDRPAGTIVVLSATDEDEGENARITTYMEDSIPOPRIDADTGAVTQAE 772  
DB 790 OSSHTVSVSEDRPVGTSTATLSANDEDEGENARITTYVLOPFPQFRIDPDSGTMYTMC 849  
QY 773 LDYEDQVSTYLAITARDNGIPOKSDPTTYLEILVNVNDNAPOFLDSYOGSVYEDVPPT 832  
DB 850 LDYENQVAYTLTIMQDNKIPOKSDTTTYLEILILANDNAPQFLMDPFOGSLFEDAPST 909  
QY 833 SVLQISATDRDSGLNGRVYTFQGGDDGDEPIVESTSGIVTTLRLRLRENAQYVLTAY 892  
DB 910 SLAQVATDRDSGPNRGLLYTFQGGDDGDFYIPTSGVIRTRQRLDRBENAVVNNMAL 969  
QY 893 AVDKMP- PARTPMEVTVVLDVNNNPVTFEODEDEVDVYEENSPIGLAVARTATDPBG 951  
DB 970 AADRSPPLPLASVSEIQTVIDINDNAPFEDDELFTVBEENNPVGSVYAKIRANDPBG 1029  
QY 952 TNAQIMYOIVBENIPEVFOJDI FSGELFALVLDYEDRPEYVLTQATSAPLVSATYHV 1011  
DB 1030 PNAQIMYOIVBGMHFRQJDLNKGDLAMVGLDEVRREYVLTQATSAPLVSATYHI 1089  
QY 1012 RLIDRNDNPVLGNFELLFNNTVYTRSSSFPFGAIGRVPADPDISLTSYSPFERGELS 1071  
DB 1090 LLDVQNDNPVPLPDEQOILFNNTVYTRKSNSPPTGVIGCI PAHDPVDSLSNTVFGVGNELR 1149  
QY 1072 LVLNAASTGELGLSALDNNRPLEAIMSVLVSVDGVHSTYAOQALAVTTITTEMLTHSITL 1131

DB 1150 LILLDPATGELQLSHDLNNRPLEALMEVSVDGHSYTAFCITLRYTITTDMLTNSITV 1209  
QY 1132 RLEDMSPRRFSLPLGLPTQAVATLATPDPHVVPVNNQORDPACGHILNLSVGGPP 1191  
DB 1210 RLNNMSQEKFPSPPLALPVEGAAVLSSTKDDVFFNNQNDTV--SSNLTNTFALLPBG 1268  
QY 1192 GPGGPPPLPSEHDLQERLYNRLSLTASAOVLPEPDNI CLAREPCENYMRCVSYLRPDS 1251  
DB 1269 GVRG--QFPSEHDLQEQYLYNRILTITITSTGRVLPFDNICLAREPCENYMRKVSYLRPDS 1326  
QY 1252 SAPPLASSSVLFRPHVYGLRCRCPPGFTGDCYCEVYDLCYSRPOGPHGRCSREGYT 1311  
DB 1327 SAPPLSSTTVLFRPHIPINGLRCRCPPGFTGDCYCEVYDLCYSRPOGPHGRCSREGYT 1386  
QY 1312 CLCRDGYTGBHCEVARSRGCRCPVCKRNGTCVNLVJGGFPKDDGSGDPREKYCOVTRS 1371  
DB 1387 CECFEDFTGEHCEVDANKSGRCANGVCKRNGTCVNLVJGGFPKDDGSGDPREKYCEVTRS 1446  
QY 1372 PPAHSFITERGLRORFHTLALSPATKERDGLLYNGRFNEKHDFVALLVIOEQVOLTFS 1431  
DB 1447 FPQGSFVTFRGLRQRFHTTISLTPATQBERNGLLYNGRFNEKHDFVALLVIOEQVOLTFS 1506  
QY 1432 AGESTTYSPTVPGSVSDQMHVYQLKTYNTPLLSGTGLPOGSEBQKAVVTVDGCDTG 1491  
DB 1507 AGEHTTVAPVPVSGSVSGRHHSVQVYNNKPNIGHGLPHGPGSEKMAVTVVDDCDTMM 1566  
QY 1492 ALRPGSVLYGNSCAAQOGGSKSLDTLGPGLLGVVDPLESPVPYRRORVCGKRNIOV 1551  
DB 1567 AVRREKDIIGNYSCAAQGYOTSXSLDTLGPPLLGGVNNLEBDPVNHRORVCGKRNLSV 1626  
QY 1552 DSRHIDMADPITANNGVVPGCPAKKNVCDNNTCHNGGTCVNOQMDAFSCCECPFGGSKCAQ 1611  
DB 1627 DGKAVDAMAGPITANNGTREGCAABNPFDCGRCCQNGGTCVNRMMNYLCECPILPFGKNCED 1686  
QY 1612 EMANPQHLGSSLYAMHGLSLPISQPYTSLMPTTROADGVLLQALTRGSTITLQREG 1671  
DB 1687 AMPHPOLFSGSVSVWSLSNLTITISVPWTGLMFRTRKEDSYLMEATSGSPSFLQILINN 1746  
QY 1672 HMLSVSEGTGLQASLSRLPERANDGPMHQAQL---GASGGGHAI-LSFPDQGOQAB 1727  
DB 1747 YLOFVSHSGPDEVSWMLSGLRVTDBGEHNLILTLKVKKEDSEMKHLYTMMTLTGMDQNK 1806  
QY 1728 GNLCPRHLGLHLSNITVGGIGPAGVYARGFRCGLQGVRSVDRPBGVNSLDPHSGBSINV 1787  
DB 1807 ADIGKMLPGLTVRSVYVVGASSEDKSVYRFRGRCMGQVRMGGTPTNNATLMMNNALKRV 1866  
QY 1788 EOGCSLDPDCDSNFCPANSYCSNDWDSYSCGCDPGYTGDNCTNVCDLNPCEHOSVCTRKP 1847  
DB 1867 KDGCDDVDPCTSSPCPNSRCHDAMBDYSVCVCKGYLGINCVDACHNPPCENMGACVRSF 1926  
QY 1848 SAPHGTCCECPNNYLGVCYCFRIDIQPCPRGMWGHPTGCPNCDVSKGPPDCNTSGECH 1907  
DB 1927 GSPGTCVCEBPSHYGYCEKLDLPCPRGMWGNPVGCPCHCAVSKGDPDCNTNNOCCQ 1966  
QY 1908 CKNHYPFGSPPTCLCDCYPTGSLSRVCPBEDQCCCKEFGVIGRQCDRCNPPAEAYTN 1967  
DB 1987 CKENYVLLAODTCLPCDCCPRHSHSRTCDMAITQCAKCGVIGRQCNRCNPPAEAYTTL 2046  
QY 1968 GCEVNTDSCRAITAGITWMPRTFRGLDPAAPCPFGSGTAVRCHDEHRCMLPPLFNCTS 2027  
DB 2047 GCEVITNGCPAFAGITWMPQTEKGPAAVPCPGKSGVGNVRHSGSGKGMVLPPLFNCTT 2106  
QY 2028 ITFSELKPARLORNSGSDSGSOOLALLLRNAOTHTAGYFSDVKKVAYQOLTRLLAH 2087  
DB 2107 ISFTDLAAMKESLKNITQVDBGARLOLVYALMSATYHTGTLPANDVTRAYQLGHHVLOH 2166  
QY 2088 ESTORGGLSATODVHETNMLLRVGSALDPTANKRHEMLIQOEGGTAMLLQHTEAYASA 2147  
DB 2167 ESMOQGFDLATQADDEHVIDHSHGSALLABATYAAWEQIORSBGGAQLRLRLEGVFSN 2226  
QY 2148 LAQNMHTTYSPTIYVTPNVIVISVRLDKGNPAGAKLPREALRGEOPDLETTVILPBS 2207

```

Db 2227 VAAVNRRTYLRPFVITVANNILAVDIFDKENFTGAVRPFDTIHEEPRELESSVSPAD 2286
Qy 2208 VFR-----ETFPVVRPAG-----PGEAQBPPEELARQRRPPELQSGAVASVITYT 2254
Db 2287 FFRPPEKEGFLRPAGRRTPPTTRPGCTERREAPISRRRRPDDAGPAAVALITYT 2346
Qy 2255 LAGLLPHNYDPDRKSLRVPKRPPIINTPVVSIIVHDEBELPRALDKPVTVQFPLETBER 2314
Db 2347 LQQLLBERYDPRDRSLRLPHRPPIINTPMVSTLYSEKAPLPRLERPVLYEFALLVEER 2406
Qy 2315 TKRICYFMNHSILVSGGWSARGCEVFPNESHVSQCCHMTSPAVLMDVSRREGEIL 2374
Db 2407 TKVVCVFMNHSILVSGGWSARGCELSNRKTHVACQCHTSFAVLMDISRREGEVL 2466
Qy 2375 PLKTLTYVALAGVTALMLTLPFLTLRLRSNOHGIRRLTALGAOLVFLGAINOD 2434
Db 2467 PLKIVTYAASVSLALLVAFVLISLVKRLSHLSIHGLAVALLPSQLVFTYGINORE 2526
Qy 2435 LPPACTVIALILHFLYLCFTSMALLBALHLYRALTEVRDVTGPMRFTYMLGVPAFTT 2494
Db 2527 NPELCTVVALILHLYIMSTFAMTLVESLHYRMLTEVRNIDTGPMRFTYVVGWIGIPALYT 2586
Qy 2495 GLAVGLDPGSGNDPCWLSIVDTLMSFAPGAPAVASVFIYIILAAASCAAGQGE 2554
Db 2587 GLAVGLDPQSGNDPCWLSLQDTLLWSFAPGIGAVIINTVTSVLSARVSCRRKHYTG 2646
Qy 2555 KKGPGVGLQSPFVALLLSATWLLALSVNSDTLLFHYLPATGNCQGPPIFYVYVLSK 2614
Db 2647 KKGIVELLRATFLILLISATWLLGLLAVNRDLASTHYLPATSGIGQFPVLLFHCVLNQ 2706
Qy 2615 EVRKALK-LACSRKPPDPALTKST-LISSYCPSPYADG-RLYQPYGDSAGSLHSTS 2670
Db 2707 EVRKHLKGVLGKRLHLEDSATTRATLLTRSLNCTTFGDPGPMRLTDLGESTASLDSIV 2766
Qy 2671 RS-----CKSQPSYIPFLIREESALNPGCGPGGLGDSGLFLEGODQ 2713
Db 2767 RDEGIQKGVSGVSLVNGSHBPASLMPRSCXK-----PPG----- 2802
Qy 2714 HDEPTDSDSLSEDQSGSYASTHSDSEEEBEEBAAFPQEGQMDSLCPGARLP 2773
Db 2803 --HDSQSDSEL-DEQSSSYASHSDSDSDGVABEK-----MDPAR-----A 2845
Qy 2774 LHSYER-DGPGPGKAPWPDFTTAKESGNGAP-----BERLRENGDAL 2818
Db 2846 VHSYTPGDAVANHVPMGPQSLAESDSBPSKPLKVTKYVSEILHREQSHGEXEP 2905
Qy 2819 SRGSGPLPGSSAOP---HKGILKKKCL---PTISEKS--SLNLPLEQCTGSSRGSS 2869
Db 2906 PDESQGAARLASQPEQRKGLKNKVTYPPPLTTEQTLKRLBKADCRQSPITSR 2965
Qy 2870 ASBGSRGAP---PPRPPRQSLQEOQLNGVPIAMSIKAGTVNDESSGSE 2915
Db 2966 TSSLSGSGPCATVYSRGRERGRDLHNGV---AMVVRIGSAGADSDS 3012

```

## RESULT 11

ABP81979  
ID ABP81979 standard; Protein; 3014 AA.

AC ABP81979;

DT 04-MAR-2003 (first entry)

DE Human GPCR CELSR1/Flamingo protein SEQ ID NO:444.

XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor; modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related disease; cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;

```

KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  

KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  

KW ulcer.  

XX Homo sapiens.  

XX WO200261087-A2.  

XX 08-AUG-2002.  

XX 19-DEC-2001; 2001WO-US50107.  

XX 19-DEC-2000; 2000US-257144P.  

XX (LIFE-) LIFESPAN BIOSCIENCES INC.  

XX Burner GC, Roush CL, Brown JP;  

XX WPI; 2003-046718/04.  

XX N-PSDB; AB242827.  

PT New isolated antigenic peptides e.g., for G protein-coupled receptors  

PT (GPCR), useful for diagnosing and designing drugs for treating  

PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  

PT cancer or autoimmune diseases  

XX PS Disclosure; Fig 1; 523pp; English.  

XX The present invention describes antigenic peptides (1) comprising:  

CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  

CC acids. Also described: (1) an assay for the detection of a particular  

CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  

CC and (2) an isolated antibody having high specificity and high affinity  

CC or avidity for a particular GPCR. (1) can be used as GPCR modulators and  

CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  

CC an antibody against a particular GPCR, and in the production of specific  

CC antibodies. The peptides and antibodies are also useful for detecting the  

CC presence or absence of corresponding GPCRs. The antigenic peptides for  

CC GPCRs and antibodies are useful for diagnosing and designing drugs for  

CC treating immune-related diseases, growth-related diseases, cell  

CC regeneration-related disease, immunological-related cell proliferative  

CC disease, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  

CC atherosclerosis, osteoporosis, cancer, cardiomyopathy, chronic and acute  

CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  

CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  

CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  

CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  

CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  

CC any other disorder in which GPCRs are involved. The antibodies may be  

CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  

CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  

CC exemplification of the present invention.  

XX SQ Sequence 3014 AA;  

Query Match 56.3%; Score 8754; DB 24; Length 3014;  

Best Local Similarity 55.8%; Pred. No. 0;  

Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  

Qy 12 TFPPLLLILLILLPPLGLGVPCRSLSGRGSSGACAPMGWLCPSASNMLTYR 71  

Db 57 TFRAPRELL-----DVGKDGRLAG-RKVSAGGRPL---PLQVRLV----- 93  

Qy 72 CRDAGTBLTGLVPHHDGLRVWCPSBAHPLPAPBEGCPMSCRLLGIGHLSPOGKTL 131  

Db 94 ASATATLASRL-----RAKTHLP-----CGARARLCGTGALIC--GALCF 133  

Qy 132 P-----EENHCKLA-----PRLRCOS-----CKLAQAPG- 155  

Db 134 PVPGGCAAGHSLAAPTLTPACRCPPRPCCRPICLPFGSVRLTLCLALRRAGA 193  

Qy 156 LRAQ-----ERSPESLSGRKRNVTAP-----QGPSPSYQAT 189

```



Db 194 VAVGALAEATAGTSPASPSPPPLPPNLPPEARAGARBARRGTSRGSLKPKPMNYOVA 253  
 QY 190 VVENOPAGTVPVSLRAIDPDEGEAGLEXTMDALPDRSNOFPISLDPVTGATTAEBEJR 249  
 Db 254 LFEENPAGTLIQIQAHTTIEGEEBVSITYMEGLPDERSGIFRIDSATGAVSTSDVR 313  
 QY 250 EYKSTHVFRTVAQDHGMPRSALATLITLVTDNDHPVFOOEKESLREMLLEVGYEVL 309  
 Db 314 EYKSTHVRKVAADVSTPPRSATVYITVVKOTNDHS PVFEOSEYERERREMLLEVGYEVL 373  
 QY 310 TYRATDGAAPPANALTYRLBESGSPSVFPIIDRSQVITRGRGPDREBVSYQUTYBA 369  
 Db 374 TIRASDRSPINANIRYRLVGAM----DVQIAMSNGSVSTRVADREBAEYQDLYEA 429  
 QY 370 SDQCGPGRSTYAAVFLSEVDNDNAPQFSEKRYVQVREDPVTCAPLRYATASDRDG 429  
 Db 430 NDQGNPGLSATATVYIEVEDENDNYPOFSEQNTVVOYPEVGLANTAVLRQAOTDRDG 489  
 QY 430 SNAVVHYSIMSGNARGQFYLDAGTALDVASPLDYETTKETLRYAODGGRPPLSNVG 489  
 Db 490 QNAAHYSILSGNVAGQFYLSGLDYNPLDFEDVQKYSLSIKAQDGGPPLINSNG 549  
 QY 490 LVTQVULDNDAPIFVSTPRQATLESYPLGYLVLAHQALADAGDNARLEYRLA---- 545  
 Db 550 VVSVQVLDVNDNEPLFVSSPFOATYLENVLGYPVVHTQAVDADSGENARLRYLVDTAS 609  
 QY 546 ---GVGH-----DEFTINNGTGISVAALDREEDVPFSGVEARHGTPLTA 592  
 Db 610 TPLGGSGSAGPKAPAPPPDPFOIHNSSGHTYCAELDREVEHYHSGVEADHDSBPMS 669  
 QY 593 SASVSVTVLDVNDNPTPTQPEYVRLNDAAVGTSVVTSVAVDSDHVSITTYQITSGNT 652  
 Db 670 STSVSITVLDVNDNPEVFTQPTYBLRINDEDAVGSVLTQAORDANSVITYQLTGNT 729  
 QY 653 RREBITQSGGGVSLAPLDYKLEROYVAVTASDGRQTAQIVVNTANTHRPVP 712  
 Db 730 RRRPALSORGGGLTLPLDYKOBQOQVLAVTASDGRSHAVLVINVTANTHRPVP 789  
 QY 713 QSSHVTVANNEDRPAGTTVLLISATDEDTGENARITYFMEBDSIPQFRIDADGATVTOAE 772  
 Db 790 QSSHVTVASDEBPPVGTISATISANDEDTGENARITYVQDPVQFRIDPSGMYTME 849  
 QY 773 LDYEDQVSTYLAITARDNGIPKXSTTYEILVNDVNDNAPQFLDSDYQSVYEDVPPT 832  
 Db 850 LDYENQVAVYTLTIMADNGIIPKSDTYTLEILILANDNAPQFLMDPYQSGSIFEDAPST 909  
 QY 833 SVLQISATDRDSGLNGRVYTTROGDDGDDFIVESTSGIVTARLRRENTAOVLAAY 892  
 Db 910 SILQVSATDRDSGPNGRLLYTFQGDGDGDYFISPTSGVIRTORRLRENTAVNLMAL 969  
 QY 893 AVDKAMP-PARTPMEVTVTVLDVNDNPPVFEQDEBDFVYEENSPIGLAVARTATDPBG 951  
 Db 970 AVDRSPPTPLASVETIQTIIIDNNAHPEKDELELFEENNPVGSVAKIRANDPBG 1029  
 QY 952 TNAQIMYQIVBENIDEPQOLDIFSEBELTALVDLYEDREYVLTQATSAPLVSAIYAV 1011  
 Db 1030 PNAQIMYQIVBEGDMHFQDLIDNGILRAMVELDEFERREYLVVQATSAPLVSAIYHI 1089  
 QY 1012 RLDDNDNPPVGNBEILLPNVYVTRSSSPFGCALGRPAHDPDISDLSITYFEEGNLS 1071  
 Db 1090 LTVNDNDNPPVLPDQILFNNAVTKNSNFPFGVIGCIPAHDPDVSISLNTYFVGNBLR 1149  
 QY 1072 LVLNASTGELKSLRALDNNRPLEAISMVLSVSDGHSVTAQCALRVTIITDEMLTHSITL 1131  
 Db 1150 LLLDPAQGELOLSDLDNNRPLEALMEVSDGHSVTAQCTLAVTIITDMLTNSITV 1209  
 QY 1132 RLDEMSPERFSLPLGLPTQAVAAITLPPDHVVVFNQRTDAPQGHILNLSVGP 1191  
 Db 1210 RLNNNSOEKFTSLPLALFVEGVAAVLSTKDDVFEFNQNDTV--SSNLANTFSALEP 1268  
 QY 1192 GPGGGPPLPSEDLERLYLNSLTATISAGRVLPDDNICLRFCENMRCVSVLARSDS 1251

Db 1269 GVRG--QFPSEBDLQEOIYLNRTLLTISTQVLPFPDDNICLRFCENMRCVSVLARSDS 1326  
 QY 1252 SAPFIASSSVLFRPIHPVGLRCRCPGFTDYCEVEVDLCYSRPGHRCRSREGYT 1311  
 Db 1327 SAPFISSTTVLFRPIHPVGLRCRCPGFTDYCEVEVDLCYSRPGHRCRSREGYT 1386  
 QY 1312 CLCRDGYTGEHCEVASRSGCTPGVCKXNGCTCVNLLVSGFKDCPSGDPEKPYCQVTRS 1371  
 Db 1387 CEFCEDFGHECEVDARSRCANGVCKNGCTCVNLLIGFHCVCPPGHEBRYCEVTRS 1446  
 QY 1372 FPAHSFIFRGLRQRFHTLALSPATKRGDGLLYNGRFRKHDPVVALEVQEOVLTF 1431  
 Db 1447 FPPQSFVTRGLRQRFHTLALSPATKRGDGLLYNGRFRKHDPVVALEVQEOVLTF 1506  
 QY 1432 AGESTTVYSPVPGVSDQGNHTVOLKYNPFLGQNTLPQSPSOKAVAVTVGCDPTG 1491  
 Db 1507 AGETTVYAPVPSGVSDBGRHSVQVQYKNGPNIHGLPHGPGSEKNAVTVDDCOTM 1566  
 QY 1492 ALRFGSVLGNYSCAAQGTQGGSKSLDTGPLLGGVBDLESPPVRMOPVGCNRLQV 1551  
 Db 1567 AVRFGKOIGNYSCAAQGTQSGKSLDTGPLLGGVBNLPEDPVRHNRQFVGCNRLSV 1626  
 QY 1552 DSRHIDMADEFIANNCTVPGCAKKNVCDNTHNGCTCVNQWDAFSCGCEPLRGKSCAQ 1611  
 Db 1637 DKNVDNAGFIANNCTVPGCAKKNVCDNTHNGCTCVNQWDAFSCGCEPLRGKSCAQ 1686  
 QY 1612 EMANPOHFLGSSLAHMGSLSPISQPWYLSLMPTRQADGVLLQAITRGRSTITQLREG 1671  
 Db 1687 AMPHPOLESBSGVSWSLNLIIISVPYLGIMFTRKEDSYLMARATSGSPRFLQILAN 1746  
 QY 1672 HVMLSVEGTGLQASLSLRLEGRANDGWHAQLA--GASGGPHAI-LSFDYGOQRAE 1727  
 Db 1747 YLOFREVSHGSPDVSVMLSGLRVTDGEWHLLILKXVKEDESMKHLVTMTLDYGMQNK 1806  
 QY 1728 GNLGPRHLGLHLSNITYGIFGPRAGVYARGRGLQGVRSVDTBEGVNSLDSHGESINV 1787  
 Db 1807 ADIGMPLGLTVRNVVVGASDEKSVYVRGFRGCMQGVGMGTPTNVAITLMMNNAKRV 1866  
 QY 1788 EOGGSLPDCDSNCPANYSYCSNDMDSYS CSCDEGYGDNCTNVCNLPCEHSGVCTKP 1847  
 Db 1867 KOGCDVDPCISSPCPNRSCHDAMBDYSCVCDGYIGINVCDAHNLPCENMAGCVASP 1926  
 QY 1848 SAPHGTCPCPNYLGPCETRIQPCPCRGWGHPTGCPNCVDYSKGFDPCKNTSGECH 1907  
 Db 1927 GSPGYVCECPSHYPCEBKLDLPCBRGMGNVPCGPHCAVASKGFPDCKNTNGCQ 1986  
 QY 1908 CKENHYRPGSPGTLCDQVPTGSLSRVCPDEBQCCPKPVITROCRCNPFABEVTN 1967  
 Db 1967 CKENYITLADQDTCPCDPCPHGSHSRCDMATQCAKCEPVIGRQCRCNPFABEVTN 2046  
 QY 1968 GCEVNYDSCPRALBAGIWMPETRFGLPAAAPCPKGSFGTAVRHCDERHGMPLPMLFNCTS 2027  
 Db 2047 GCEVITNGCPRAFBAGIWMPETRFGLPAAAPCPKGSFGTAVRHCDERHGMPLPMLFNCTS 2106  
 QY 2028 ITSEBLKFAERLORNSGLDSGRSQQLALLRNATQHTAGFSGDYKAVAYQATRLIAH 2087  
 Db 2107 ISFVDLAMEKISRNETQVNDGARALQIVRALRSATQHTGTLFQNDVYATAYQLGHTVLOH 2166  
 QY 2088 ESTORGFLASATQVHFTENLTVRSALDTPANKRHEMLQOTEGGTAMLOHAYASA 2147  
 Db 2167 ESMOQGFDLAQTQADPHBEDVTHSGSALLAPATMAHQIORSRGGTQQLRLREGFSN 2226  
 QY 2148 LAQNMRTYLSPPFTIVPNIVISVVRLDKGNFAGAKLPRYEALRGEPDLETTVILPES 2207  
 Db 2227 VARVVRSTYLRPFVITVANMLAVDIDKFNFTARVPDPTIHBFPRLBESSVSPAD 2286  
 QY 2208 VFR----ETPPVVRPAG-----PBAQEPBELARQRRHRLBSGEAVASYIYRT 2254  
 Db 2287 FFRPPEKEGELRPAKRRTTPQTRBPGTEREAPISRRRRHDDGQFVAVALIYRT 2346  
 QY 2255 IAGLLPHNYDPPKSLVYKXPIINTPVVSSIVYDDEBELPRALDKPVTQVFRLLFTEER 2314  
 Db 2347 LGQILPERYDPRKSLRPLRPIINTPVVSTLVISGAPLPRPLERFVLVFAALLBYER 2406



QY 302 LEVGEVUTVATDGDAPNANILYRLLEGSGSGSEVFEIDPNSGVITRTGPDVDRREEV 361  
 Db 366 LEVGEVUTVATDGDAPNANILYRLLEGSGSGSEVFEIDPNSGVITRTGPDVDRREEV 421  
 QY 362 SYOLIVASDQGRDPRSTTAAVFLSYEDDNDNAPQSEKRYVYQVVEDTPGAPVARY 421  
 Db 422 EYOLLVANDQGRNRPISATRTVYIEVEDENDYPOQSEKRYVYQVVEDTPGAPVARY 481  
 QY 422 TASDRKSNVWHSYINSNGARQGYLDAQDGLDVVPLDYETTKETYLVRADQGR 481  
 Db 482 QATRDQONNAIHYSILSGNVAQGYLHSLGILDVNPDLDFEDVQKYSISIKADQGR 541  
 QY 482 PPLSVSGUYYVQVLDINDNAPIFSTTPQATVLESVPLGLVLAHQVLDADQDNARLE 541  
 Db 542 PPLSVSGUYYVQVLDINDNAPIFSTTPQATVLESVPLGLVLAHQVLDADQDNARLE 601  
 QY 542 YRLA-----GVGH-----DPPFTINGNWTSVAAELDRREVDPYSGVEARD 584  
 Db 602 YRLVDTASTPLGGSGAGKRNAPPTDPFPQIHNSGMLTVOCHELDRESEVHSGVEARD 661  
 QY 585 HGTPALTAASVSVTVLDVNDNPPFTPTPEYTVRLMEDAVGTSVTVYSAVDRAHAYIT 644  
 Db 662 HGSPFMSSTSVSLTVLDVNDNDPFTQPTBELRLMEDAVGSSVTLTQADRDANSVIT 721  
 QY 645 YQITSNGTRNRPSTISQSGGLVSLALPLDYKLERQVYLAATASDGTRODTAQVANTD 704  
 Db 722 YQITSGNTRNRPSTISQSGGLVSLALPLDYKLERQVYLAATASDGTRODTAQVANTD 781  
 QY 705 ANTRRPVQSHYTVANVEDRPACTTVLISATBDEGENARITYFEMDSIPORINDAT 764  
 Db 782 ANTRRPVQSHYTVANVEDRPACTTVLISATBDEGENARITYFEMDSIPORINDAT 841  
 QY 765 GAVTTQALDEDOVSYTLATITARDNGIPKSDTYTLEILVNDVNDNAPQFLRDSYQGSV 824  
 Db 842 GAVTTQALDEDOVSYTLATITARDNGIPKSDTYTLEILVNDVNDNAPQFLRDSYQGSV 901  
 QY 825 YEDVPFPTSVYQIATDDBDGLNGRVFTFGQGDGDGPVYESTSGIVTLRLDRNV 884  
 Db 902 YEDVPFPTSVYQIATDDBDGLNGRVFTFGQGDGDGPVYESTSGIVTLRLDRNV 961  
 QY 885 AQVYLAAYVDKAMP-PARTPMEVTVTVLDVNDNPPVEODEPDPVFEVENSPIGLAVARV 943  
 Db 962 AQVYLAAYVDKAMP-PARTPMEVTVTVLDVNDNPPVEODEPDPVFEVENSPIGLAVARV 1021  
 QY 944 TATPDEGTNAQIMYQIVEGNIPEVQOLDIFSGELTALVDDYEDREYVLYQATSAPL 1003  
 Db 1022 TATPDEGTNAQIMYQIVEGNIPEVQOLDIFSGELTALVDDYEDREYVLYQATSAPL 1081  
 QY 1004 VSRATVHRLDRNDNPPVLGNFELFNNTYTNSSSPFGAIGRVPHNDPDISDLYTS 1063  
 Db 1082 VSRATVHRLDRNDNPPVLGNFELFNNTYTNSSSPFGAIGRVPHNDPDISDLYTS 1141  
 QY 1064 FERGENELSLVLLNASTGELKLSRALDNNRPLEAIVSVLV--DGVASVTAOCALRTVITTD 1122  
 Db 1142 FERGENELSLVLLNASTGELKLSRALDNNRPLEAIVSVLV--DGVASVTAOCALRTVITTD 1201  
 QY 1123 EMLTHSITLRLDMSPERFISPLGLFTQAVAAATLATPPDHVVVFNVOJTDABGHILN 1182  
 Db 1202 EMLTHSITLRLDMSPERFISPLGLFTQAVAAATLATPPDHVVVFNVOJTDABGHILN 1260  
 QY 1183 VSLSVGPFGGPGGPPPLPSDLOKRLVANSILTAISQVLPDNDNICARECEYMR 1242  
 Db 1261 VSLSVGPFGGPGGPPPLPSDLOKRLVANSILTAISQVLPDNDNICARECEYMR 1318  
 QY 1243 CVSVLRPDSAPFLASSSVLFRPIHPVGLRCRCPGFTGDCETEVLDCTSRCPGPHGR 1302  
 Db 1319 CVSVLRPDSAPFLASSSVLFRPIHPVGLRCRCPGFTGDCETEVLDCTSRCPGPHGR 1378  
 QY 1303 CRSSBGGYTCLCRDGTGEHCEVSARSGRCTPGVCNKGCTCVNLLVSGFKDCPSGDEFK 1362  
 Db 1379 CRSSBGGYTCLCRDGTGEHCEVSARSGRCTPGVCNKGCTCVNLLVSGFKDCPSGDEFK 1438  
 QY 1363 PYCQVTRSPFAHSITFRGLRQRHFTLALSPATKERDGLLVNGRNMGHDVVALBEV 1422

Db 1439 PYCQVTRSPFAHSITFRGLRQRHFTLALSPATKERDGLLVNGRNMGHDVVALBEV 1498  
 QY 1423 QEOVQUTFS--AGSTTVSPVAVGVSQDGMQVOLKTYNK-----PLGQT 1468  
 Db 1499 QEOVQUTFSAGAGETTITVAPKPSGVSQDGMQVOLKTYNK-----PLGQT 1558  
 QY 1469 GLPQSPBORAVVTVVDCDTGVALRPGSVLGVNSCAAGSTQGGSKSLDTGFLLGCV 1528  
 Db 1559 GLPQSPBORAVVTVVDCDTGVALRPGSVLGVNSCAAGSTQGGSKSLDTGFLLGCV 1618  
 QY 1529 PDLBESFPVNRQPVGCMRLQVDSRIIDMADPLANNGTVPCCAKOVCDNSTCHNGCT 1588  
 Db 1619 PDLBESFPVNRQPVGCMRLQVDSRIIDMADPLANNGTVPCCAKOVCDNSTCHNGCT 1678  
 QY 1589 CVNOMDAPSCPCPGFSGCAOEMANPOHFLSSLAHMGSLPISQWYLSIMPTRO 1648  
 Db 1679 CVNOMDAPSCPCPGFSGCAOEMANPOHFLSSLAHMGSLPISQWYLSIMPTRO 1738  
 QY 1649 ADGVLLQATIRGRSTITLQREBHVMLSVETGLOASLRLBGRANDGMHNAQLA-- 1706  
 Db 1739 EDSVLMATSGGPTSPRLQILANNVLOFEVGHGSDVSVMLSGLRVTDGEWHLLIELON 1798  
 QY 1707 -GASGGEHAL-LSFDYGOORABGNLGRHLGHLSTYTGGLPGRPGYARGRGLG 1764  
 Db 1799 VKDESEKHLVTWTLTDGMDQNRADIGMLPGLTVSVVVGASBDRKSVRGRGCMOG 1858  
 QY 1765 VRVSDTREGVNSLDPSSHSINVEGCSLPDCDSNCPANSVCSNDMSVSGSCDPCGY 1824  
 Db 1859 VRVSDTREGVNSLDPSSHSINVEGCSLPDCDSNCPANSVCSNDMSVSGSCDPCGY 1918  
 QY 1825 GDNCTVCDLNPCEHOSVCTRKPSAPHGTYCEBPNTAGYCESTRIDQPCPRGMGHTC 1884  
 Db 1919 GDNCTVCDLNPCEHOSVCTRKPSAPHGTYCEBPNTAGYCESTRIDQPCPRGMGHTC 1978  
 QY 1885 GPCKCDVSKFPDPCNKTSGECHKENHRRPQSPPTLDCYPTGSLSRVCPBEDQCP 1944  
 Db 1979 GPCKCDVSKFPDPCNKTSGECHKENHRRPQSPPTLDCYPTGSLSRVCPBEDQCP 2038  
 QY 1945 CKPGVIGRQCDNCPNPAEVTNGCEVNYDSCPRATEAGIMPTRETRGLPAAAPCPKGSF 2004  
 Db 2039 CKPGVIGRQCDNCPNPAEVTNGCEVNYDSCPRATEAGIMPTRETRGLPAAAPCPKGSF 2098  
 QY 2005 GTAVRHCHERGMPLPPLFNCTISITPSBELKGFARLQRNESGLDSGSQOLALLRNATQ 2064  
 Db 2099 GTAVRHCHERGMPLPPLFNCTISITPSBELKGFARLQRNESGLDSGSQOLALLRNATQ 2158  
 QY 2065 HTAGVSGDVKAVYQVATLRLAHESVQRGSLSATQDVHTEMLNARGLALDTANRBN 2124  
 Db 2159 HTAGVSGDVKAVYQVATLRLAHESVQRGSLSATQDVHTEMLNARGLALDTANRBN 2218  
 QY 2125 ELIQTEGTAMVLLQHYEAVASALQNMRTYLSPTITVPTNIVISVRLDKGFAGAKL 2184  
 Db 2219 ELIQTEGTAMVLLQHYEAVASALQNMRTYLSPTITVPTNIVISVRLDKGFAGAKL 2278  
 QY 2219 EOTORSBGGRNOLRLRLEGFSVNAVARNRTYLRPIVYANVLAVIDDKENFTGARV 2278  
 Db 2279 EOTORSBGGRNOLRLRLEGFSVNAVARNRTYLRPIVYANVLAVIDDKENFTGARV 2338  
 QY 2232 RQRBHPELSQSGRAVSVIYRITLAGLPHANVOPDRKSLVPPRPITNTEVSVISVUDE 2291  
 Db 2339 RQRBHPELSQSGRAVSVIYRITLAGLPHANVOPDRKSLVPPRPITNTEVSVISVUDE 2398  
 QY 2292 ELPLALDKKPVYQFRLLETERTKDICYFWMNHSILVSGTGSAGACVAVFNESHVSC 2351  
 Db 2399 ELPLALDKKPVYQFRLLETERTKDICYFWMNHSILVSGTGSAGACVAVFNESHVSC 2458  
 QY 2352 QCNHMTSPAVLMVSRRENCEIIPKLTLYVALGVTLAALLPFPPLTLRLIRSNQGI 2411  
 Db 2459 QCNHMTSPAVLMVSRRENCEIIPKLTLYVALGVTLAALLPFPPLTLRLIRSNQGI 2518  
 QY 2412 RRMVTAALGAQVFLGIGNQADLPACTVIALPLPVYCTESMALLBAMLYRALTEV 2471

```

Db 2519 HKHIAVALFLSOLVFINQNTENPLCTVVALIILHYINSTEAWTLVESILHYRMILTEV 2578
Qy 2472 RDVNTGMRFYMYLGMVGPATITGLAVGLDPEGVGNDFPCMLSTYDTLWSPAGPAFAV 2531
Db 2579 RNIIDTGMRTTYVVGWICIPALVTGLAVGLDPOGIVGNDFPCMLSTYDTLWSPAGPAFAV 2538
Qy 2532 SMSVFLYIILAAASCAAROGFEKKGPVSGIOPSFAVILLISATWLLALISVSDTLLEH 2591
Db 2639 IIMVTSVLSAKVSCQKHYYGKGIIVSLRTAFILILLISATWLLGLAVNRDLSPH 2698
Qy 2592 YLPAITCICIGPPFLISYVLSKEVKALK-LACSRPSPDPALTKST-LTSSVNCPS 2649
Db 2699 YLPAISGIGPPFLVLPCHVINOERKHLKGVLGKRLHEDSATYTRATILTLISLNCMT 2758
Qy 2650 YADG--RLYOPYGDSAGSLHSTRS-----GKQPSYIPLLRBSAL 2690
Db 2759 FCGOPDLRTDLGSESTSLDSIYRDEGIQKLGVSGLVRSKRGPPDASLMPRCKD---- 2814
Qy 2691 NPGQPPGLGDPGSLFLEGDDQHDPDTSDSDLSLEDDSGSYASTHSSDSSEEEEEE 2750
Db 2815 ----PPG-----HDSDBSELST-DEQSSSYASSHSDSDSDDDGVGAE 2851
Qy 2751 EEAAPFEGQWDSILSGAERLPLHSTPK-DGGGPGKAWPDPGTAKSSGNAP-- 2807
Db 2852 EK-----WDPARG-----AVHSTPKGDVAHVPAQMPQSLAESDSEDPGKPR 2897
Qy 2808 -----BERLRENGDALSRGSLGPLPGSSAQP--HKGILKKCL----PTISE 2849
Db 2898 KVEITKSVELHREHSGSHREYPPDQESGAARLASQPEQOSILKMKYTPPLTLITE 2957
Qy 2850 KS--SLIRLPLEOCTGSSRSSASSESGRGP---PPRPPROSLQOLNGVMPIMSIK 2903
Db 2958 QTLKGRIRKRLADCEQPTSSRTSLSGGPDCAITVKSPEGRGRDLHNGV---AMNVR 3014
Qy 2904 AGTVDDSSGSE 2915
Db 3015 TGSADGSDSE 3026

RESULT 13
AAW27161
ID AAW27161 standard; Protein; 2707 AA.
XX
AC AAW27161;
XX
DT 08-DEC-1997 (first entry)
XX
DE Mouse receptor ME2.
XX
XX Chemical screening; epidermal growth factor; EGF; repeat; cadherin;
KW neuronal cell division; differentiation; cell physiology; trauma;
XX neurodegenerative disease; adenylate cyclase; CAMP.
XX
OS Mus musculus.
XX
XX Key Location/Qualifiers
XX FH 1..132
XX Region /label= CD1
XX FT /note= "Cadherin-like repeat"
XX FT 133..239
XX Region /label= CD2
XX FT /note= "Cadherin-like repeat"
XX FT 240..349
XX Region /label= CD3
XX FT /note= "Cadherin-like repeat"
XX FT 350..466
XX Region /label= CD4
XX FT /note= "Cadherin-like repeat"
XX FT 467..568
XX Region /label= CD5
XX FT /note= "Cadherin-like repeat"
XX FT 569..671
XX Region /label= CD6

```

```

FT /note= "Cadherin-like repeat"
FT 672..778
FT Region /label= CD7
FT /note= "Cadherin-like repeat"
FT 779..883
FT Region /label= CD8
FT /note= "Cadherin-like repeat"
FT 884..989
FT Region /label= CD9
FT /note= "Cadherin-like repeat"
FT 1128..1145
FT Region /label= Div EGF
FT /note= "Divergent epidermal growth factor-like repeat"
FT 1146..1178
FT Region /label= EGF1
FT /note= "Epidermal growth factor-like repeat"
FT 1186..1219
FT Region /label= EGF2
FT /note= "Epidermal growth factor-like repeat"
FT 1432..1463
FT Region /label= EGF3
FT /note= "Epidermal growth factor-like repeat"
FT 1655..1686
FT Region /label= EGF4
FT /note= "Epidermal growth factor-like repeat"
FT 1690..1724
FT Region /label= EGF5
FT /note= "Epidermal growth factor-like repeat"
FT 1773..1809
FT Region /label= DIVERGED_EGF
FT /note= "Divergent epidermal growth factor-like repeat"
FT 2244..2270
FT Region /label= TM1
FT /note= "Transmembrane portion"
FT 2282..2295
FT Region /label= TM2
FT /note= "Transmembrane portion"
FT 2309..2333
FT Region /label= TM3
FT /note= "Transmembrane portion"
FT 2346..2364
FT Region /label= TM4
FT /note= "Transmembrane portion"
FT 2389..2407
FT Region /label= TM5
FT /note= "Transmembrane portion"
FT 2432..2452
FT Region /label= TM6
FT /note= "Transmembrane portion"
FT 2463..2484
FT Region /label= TM7
FT /note= "Transmembrane portion"
XX
XX MO9707209-A2.
XX
XX 27-FEB-1997.
XX
XX 14-AUG-1996; 96WO-GB01980.
XX
XX 21-AUG-1995; 95GB-0017098.
XX
XX (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX Hadjantonakis A, Little P;
XX WPI: 1997-165293/15.
XX DR N-PSDB; AAT85320.
XX
XX Screening chemicals which interact with receptor contg. EGF-like
XX and/or cadherin-like repeats - to develop pharmaceuticals for
XX treating neurodegeneration, trauma, etc.
XX
XX Claim 10; Fig 10; 55pp; English.

```

XX A method has been developed for the screening of a chemical for  
 CC subsequent use as a pharmaceutical agent. The method involves contacting  
 CC the chemical with a receptor comprising epidermal growth factor  
 CC (Egf)-like repeats and/or cadherin-like repeats and determining whether  
 CC the chemical interacts with the receptor to form a complex. The present  
 CC sequence represents the mouse receptor ME2, which can be used in the  
 CC above method. The method is used to screen for chemicals which can cause  
 CC neural trauma, treat neurodegenerative disease, affect cellular physiology,  
 CC cyclase production, increase cAMP levels and promote neural growth.

Sequence 2707 AA:

Query Match 54.6%; Score 8495; DB 18; Length 2707;  
 Best Local Similarity 60.3%; Pred. No. 0;  
 Matches 1601; Conservative 386; Mismatches 569; Indels 98; Gaps 20;

192 ENQPGTPTASLRAPDIDPDEGRLEMTMDALFDSNQPSLIDPTGAVTTAEBLDRRT 251  
 35 EKRKGTAVIEIRAPDIDPDEGRLEMTMDALFDSNQPSLIDPTGAVTTAEBLDRRT 94

252 KSTHVRVTAODHGMFRSALATLTILVTNDHDPVFEQEKESLRNLEFVGEVLTAV 311  
 95 KOTHLAKSAVDHGSFRSALATLTIVTSDINDHSPVFEQSEKIRKLENGVIEVLT 154

312 RATDGDAPPNANILYRLLESGSGSPSEVEIDPRSGVIRTRGPVDRBEVESYQLTVEASD 371  
 155 RATDGDAPPNANMRVYRLLEGAGG---VPEIDARSGSVTRAVVDREBAEYQLVLEAND 210

372 QORDGPRSTTAAPLVSEDDNDNAQSEKRYVQVREDVTPGAVLRYTASDDEKSN 431  
 211 QQRNPEPLSASATVHIVVEDENDNPQFSEKRYVQVREDVTPGAVLRYTASDDEKSN 270

432 AVVHYSINGNARGOFLYDAOTGALDVSPFLDYETKEKTELRVRAQDGRPPLSNVGLV 491  
 271 AAHHSIVSGNKGOFYLSLGSGLDVINPLDFEALRETTIKAKDGRPPLINSGLV 330

492 TVQVLDINDNAPIFVSTPFOATVLESVPVLYVLAQALDADAGNARLEVLAGVH-- 549  
 331 SVQVLDVNDNAPIFVSSPFOAVLENVPLGHSVLMHQAVDADAGENARLYQLVDTASTI 390

550 -----DPEPTINGCMISVAALDREEDVDFYSPGYEARDHGTPLATASA 594  
 391 VCGSSVDSENPAADPFPQIHNSGMITVCAELREVEHYSPGEAVDHSPAMSSSA 450

595 SVSVTVLDVNDNAPFTQPEYTVRLNEDAAVGSVTVYSAVDRDASHVTVYITSGNTRN 654  
 451 SVSITVLDVNDNAPFTQPEYTVRLNEDAAVGSVTVYSAVDRDASHVTVYITSGNTRN 510

655 RPSITSQSGGLVSLALPLDYKLERQYLAVTASDGTRODTAQIVVNTDANTHRPVPOS 714  
 511 RPLSSQSGGLITLALPLDYKQEROYLAVTASDGTRODTAQIVVNTDANTHRPVPOS 570

715 SHYTVNVDNRAGTIVTLISATDEPTGEMARITPMEDSIQOFRIADPTGAVTTQAEID 774  
 571 SHYTVSVSDRVPVGSIAITISATDEPTGEMARITVYLEDVPVQFRIDPTGAVTTQAEID 630

775 YEDQVSYTLAITARDNGIPQKSDTTYLTILVNDVNDNAPQPLRDSYQGSVYEDVPEPTSV 834  
 631 YEDQAAVTLAITARDNGIPQKSDTTYLTILVNDVNDNAPQPLRDSYQGSVYEDVPEPTSV 690

835 LQISATDRDSGLNGRVFTYFGGDDGDDGFVISTSGIVRTLRRLDRENVQYVLRAYAV 894  
 691 LQVSTDRDSGLNGRVFTYFGGDDGDDGFVISTSGIVRTLRRLDRENVQYVLRAYAV 750

895 DKGMP-PARTPMEVTVTVLVNDNAPVFEQDEPDVPEVENSPIGLAVARVATDDEGN 953  
 751 DKGSPPPLSASVGIQVSVLDINDNPPVEKDELELFEVENSPIGLAVARVATDDEGN 810

954 AQIMTQIVEGNIPFVQDLDFSGELTALVDLDYEDRPEVLYQNTSAFLVRATVHRL 1013  
 811 AQITQIVEGNIPFVQDLDFSGELTALVDLDYEDRPEVLYQNTSAFLVRATVHRL 870

1014 LDRNDNPVLGNFELFNANNVYTNSSSPFGALIGRPADHDPIDSLTYSPFERNELSLV 1073  
 871 LDQNDNPPELPDQFLFNANNVYTNSSSPFGALIGRPADHDPIDSLTYSPFERNELSLV 930

1074 LMANSTGELKSRALDNNRPLEBAIVSVSGVSHVTAQALRVTTITDEMUTHSITRL 1133  
 931 LMDPTGELQSRDLNNRPLEBAIVSVSGVSHVTAQALRVTTITDEMUTHSITRL 990

1134 EDMSEERFLSPILGFLQVAAVATLPPDHVVVFNVOGDTAPGHIINLNVSLSVQPPGP 1193  
 991 EDMSEERFLSPILGFLQVAAVATLPPDHVVVFNVOGDTAPGHIINLNVSLSVQPPGP 1049

1194 GGGPPELPSEDLQERLYLNRSILTAISAQVLPFDNLCLEPCEYNNRCVSLAFDSSA 1253  
 1050 RG-RFPSEDLQERLYLNRSILTAISAQVLPFDNLCLEPCEYNNRCVSLAFDSSA 1107

1254 PFIASSVLPFPIHPVGLRCRCPGFTGDCYCEBVDLCYRPPCGPHRCRSREGYTCL 1313  
 1108 PFISSVLPFPIHPITGLRCRCPGFTGDCYCEBVDLCYRPPCGPHRCRSREGYTCE 1167

1314 CRDGYTGHCVSARSGRCTPGCKNGGTCVNLVVGPKDCDPSGDFEKPVCQVYTSFP 1373  
 1168 CFEDPTGSHCOVNRSGKASGVCKNGGTCVNLVVGPKDCDPSGDFEKPVCQVYTSFP 1227

1374 AHSPTTFRGLRORHFTLASFATKRDGLLYNGRFNEKADFYALFVYIQVQLTFSAG 1433  
 1228 POSFTVFRGLRORHFTLASFATQDRNALLYNGRFNEKADFYALFVYIQVQLTFSAG 1287

1434 ESTTVSPFVPGVSDGQMTVOLKTYNKPILGOTGLPQGSBOKVAVVYTDGCDYVAL 1493  
 1288 ETTTVTPVQVGVSDGQMTVOLKTYNKPILGOTGLPQGSBOKVAVVYTDGCDYVAL 1347

1494 RFGSVLGNVCAAGTQSGSKSLDLTGPIILGGVPLPESFPFRMQOFVCKMNLQVDS 1553  
 1348 HFGSVLGNVCAAGTQSGSKSLDLTGPIILGGVPLPESFPFRMQOFVCKMNLQVDS 1407

1554 RHIDMAPIANGTVPGCPACKVNCDSNTCHNGGTCVNMQMAFSCPCPLGFGKSCAQEM 1613  
 1408 RIVDMAPIANGTVRAGASGRNFCDSNTCHNGGTCVNMQMAFSCPCPLGFGKSCAQEM 1467

1614 ANPQHFLGSLVAMHGLSLPISQPMYLSMERTQADGVLOALTGRSITTLQLRSHV 1673  
 1468 PNPQHFLGSLVAMHGLSLPISQPMYLSMERTQADGVLOALTGRSITTLQLRSHV 1527

1674 MLSEVGTGLQASSLRBERGRANDGMHQAOLALASGSGPH-----MISPDYQGRSEG 1728  
 1528 RFEVSYGSDVASHQLSKSLTDDGMMHLIEL-RSAKRGKDIYTLAMTLDYGDOSTV 1586

1729 NLGPRHLGHLNITTVGGIPGPAGVARGFRGLQGVRSPTPEGVNSLDPHSGESINVE 1788  
 1587 QIQNGQLPGLKRTIVIGGVTEADKYVHRHGFRCGQGVARMGSSNTIATLNNMDLAKRYK 1646

1789 QGCSLPDPCDSNPPCANSYCSNMDYS CSCDPQYTGDNCTNVCLDLPCHESQVCTRKPS 1848  
 1647 DGCVDPEPCASSPCPPHPCGDTWDSYSCIDRGYLRKKCYDACLINPCKHVSILCALPN 1706

1849 APHGTCPCPNYLYGPCETRIIDQPCPRGMWGHFTGCGCNCDVKGPRDPCNKTSGBCHC 1908  
 1707 TPRGYSCEGFGHGGQICENKVDLPCEKGMWGNCAVPVYLSAKALIPYAT----- 1758

1909 KENHYRPGSPYCLLCDCYP-----TGSLSRVCDPEDEGQCPCKPVGIRQCRCRD 1958  
 1759 -----RPMASARRITTSPPORIVAFPVTVSPRSHSRACDMDTGCAKPGVIGRQCRCRD 1813

1959 NPFAVTTNGCVVYNDSPRALIENGIMWPRTRFELPAAAPCPKPSFGTAVAHCEHGMW 2018  
 1814 NPFAVTSLSGEVLYNCGCPRAFEAGIWPQMKFQPAVYICPKPSYVANAARHCGEKGWL 1873

2019 PPNLFCNTSITFSELSKFAERLQENESGLDGSQOALLLRNATORTAGFGSGDVAVY 2078  
 1874 PPELFCNTSISFVLDKALNKEKLNWNRMDGNBSRLAKALRNATQGNSTLFGNDVATAY 1933

QY 2079 QLATRLAHSTORFGLSATODVHTENTLAVGSLALDTANKRMWELIQOTEGTAMLL 2138  
 DB 1934 QLAARLQHRSROGPDLATREANFEDVHTGSLALPATASWGLQKRSAGAAQLL 1993  
 QY 2139 QHTBAVASALQNMRTHTLSPFTIVTNIVISVVRDLKGNFAGAKLPRYEALRGBOPLD 2198  
 DB 1994 RHEFAVFSNVARNVKTYLRPFVITVPMMLAVDIPDKLNFQAQVPRFEDIGELPREL 2053  
 QY 2199 ETTVILPESVP-----RTPVVR-----PAGEBAQEPBELARQRHRELSCGEA 2245  
 DB 2054 ESSVSPADTKEPEKKEGFPVRLTNRRTPPLTAQEPRAERETSSRRRRHDEBQORA 2113  
 QY 2246 VASVITRTLAGLPHNYDPPKSLRVKRPPIINTPVASVYDDEBELPRALDKPVTVQ 2305  
 DB 2114 VALVITRTGQLPHERYDDPHSLRPNRPVINTPVASVYDDEBELPRALDKPVTVQ 2173  
 QY 2306 FLLEETERTKPICVPMNHSILVSGTGSARCEVFNESHVSQCQNMHTSPAVLMDV 2365  
 DB 2174 FSLLEETERTKPICVPMNHSILVSGTGSARCEVFNESHVSQCQNMHTSPAVLMDV 2233  
 QY 2366 SRENEBELPLKTLTVAGVTLAALLTFEFLTLRLASNGHGRNLTALGLAQV 2425  
 DB 2234 SRENEBELPLKTLTVAGVTLAALLTFEFLTLRLASNGHGRNLTALGLAQV 2293  
 QY 2426 FLGINQADLPACTVVAIHLHFLYCTFSMALLLEALHLTRALTEVADVNTGEMRYML 2485  
 DB 2294 FMVGINQADLPACTVVAIHLHFLYCTFSMALLLEALHLTRALTEVADVNTGEMRYML 2353  
 QY 2486 GWCVPVAFITGLAVGLDPEGVGNPFCMLSTYDTLINSFAGVPAFVMSVFTYTLAARAS 2545  
 DB 2354 GWCVPVAFITGLAVGLDPEGVGNPFCMLSTYDTLINSFAGVPAFVMSVFTYTLAARAS 2413  
 QY 2546 CANOQGFKEKGVSGLOPFAVILLASATWLLASVNSDTLLFHTLPTACNCTGCPPI 2605  
 DB 2414 CQKHHYERKGVSGLOPFAVILLASATWLLASVNSDTLLFHTLPTACNCTGCPPI 2473  
 QY 2606 FLGYVVLISKERVALK-LACSRKSPDPALTYKST-LTSSYCNPSYADRLTYQPY--GD 2661  
 DB 2474 LLYCYANREVRKHLRAVLAKGLQDDSATTRATLLTRLNCCNMNTYSESSRRAHPRPQ 2533  
 QY 2662 SASGLHSTSGSGSQSYTYFLLRBSALNPGGPPGLDPSGLFL-EGDQDQHDPTMS 2720  
 DB 2534 STSLDSTTRDEGVQ-----KLVSAGPARGNHGPDPASFTPRNSKKAHGSDS 2583  
 QY 2721 DSDLSIEDDQSGSYASTHSDESEEBEERBAAPFEGQMDSLGPGARLRLHSTPK- 2779  
 DB 2584 DSELISL-DEHSSSYASTHSDESDGGEADK-----MNPAGGPA-----HSTPKA 2628  
 QY 2780 DGGPGGPKAPWPD 2793  
 DB 2629 DALANHVPAWMPD 2642  
 RESULT 14  
 ABP53570  
 ID ABP53570 standard; Protein: 4115 AA.  
 AC ABP53570:  
 XX 17-DEC-2002 (first entry)  
 DE Human NOV2 protein SEQ ID NO.4.  
 XX  
 KM Human; NOVX: cytosolic; neuroprotective; anticonvulsant; cardiovascular;  
 KM cerebroprotective; nootropic; antidiabetic; antiinflammatory; fungicide;  
 KM antineumatic; antitachytic; immunosuppressive; antiallergic; virucide;  
 KM antineumatic; antibacterial; protozoacide; antihelminthic; gene therapy;  
 KM cancer; leukaemia; lymphoma; melanoma; neurological disorder; epilepsy;  
 KM stroke; ischaemic cerebrovascular disease; Alzheimer's disease; allergy;  
 KM Pick's disease; vesicular transport disease; cystic fibrosis; goltre;  
 KM diabetes mellitus; Grave's disease; gastrointestinal disorder; vaccine;  
 KM ulcerative colitis; gastric disorder; duodenal disorder; infection;  
 KM autoimmune disease; allergic reaction; autoimmune haemolytic anaemia;  
 KM

KW rheumatoid arthritis; chromosome 3q21.3-4.  
 XX Homo sapiens.  
 OS WO200262999-A2.  
 XX  
 PN 15-AUG-2002.  
 PD  
 XX  
 PF 31-DEC-2001; 2001WO-US49976.  
 XX  
 XX 29-DEC-2000; 2000US-258928P.  
 PR 02-JAN-2001; 2001US-259415P.  
 PR 04-JAN-2001; 2001US-259785P.  
 PR 20-FEB-2001; 2001US-269814P.  
 PR 09-MAR-2001; 2001US-279863P.  
 PR 29-MAR-2001; 2001US-279832P.  
 PR 29-MAR-2001; 2001US-279833P.  
 PR 13-APR-2001; 2001US-283889P.  
 PR 18-APR-2001; 2001US-284447P.  
 PR 25-APR-2001; 2001US-286683P.  
 PR 29-MAY-2001; 2001US-294080P.  
 PR 16-AUG-2001; 2001US-312815P.  
 PR 17-AUG-2001; 2001US-313325P.  
 PR 17-SEP-2001; 2001US-322699P.  
 PR 26-NOV-2001; 2001US-333350P.  
 XX  
 PA (CUBA-) CUBAGEN CORP.  
 XX  
 PI Spytek KA, Li L, Wolenc AR, Vernet CM, Eilen A, Liu X;  
 PI Malpankar U, Shinkets RA, Tchernev VT, Spaderna SK, Gorman L;  
 PI Kekuda R, Paturajan M, Gusev V, Gangoli EA, Guo X, Shenoy S;  
 PI Raselli L, Casman SJ, Boldog F, Burgess CE, Edinger S;  
 PI Ellerman K, Gunther E, Smithson G, Miller I, MacDougall JR;  
 XX  
 XX WPI; 2002-732706/79.  
 DR N-PSDB; ABQ82327.  
 DR  
 XX New NOVX polypeptides and polynucleotides useful for treating  
 PT NOVX-associated disorders, such as cancers, neurological disorders,  
 PT disorders of vesicular transport, gastrointestinal disorders, and  
 PT autoimmune diseases -  
 XX  
 PS Claim 1; Page 17-18; 444dp; English.  
 XX  
 CC The present invention describes novel human proteins designated NOVX,  
 CC where X is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,  
 CC cytosolic, anticonvulsant, cerebroprotective, nootropic, cardiovascular,  
 CC antidiabetic, antiinflammatory, antineumatic, antitachytic, virucide,  
 CC immunosuppressive, antiallergic, antineumatic, antibacterial, fungicide,  
 CC protozoacide and antihelminthic activities, and can be used in gene  
 CC therapy. The NOVX proteins, nucleotides or antibodies can be used in the  
 CC manufacture of a medicament for treating a syndrome associated with a  
 CC human disease selected from NOVX-associated disorder, such as cancers  
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,  
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,  
 CC ischaemic cerebrovascular disease, Alzheimer's disease or Pick's  
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,  
 CC diabetes mellitus, Grave's disease, or goltre), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),  
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic  
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic  
 CC and protozoal infections. The NOVX proteins can be used as immunogens to  
 CC produce antibodies and as vaccines. The NOVX nucleotide sequences may  
 CC be used in chromosome mapping, identifying individuals from minute  
 CC biological samples (tissue typing), and in forensic identification of a  
 CC biological sample. The present sequence represents human NOV2, which is  
 CC located on chromosome 3q21.3-4.  
 CC  
 XX Sequence 4115 AA;  
 SQ  
 Query Match 50.3%; Score 7816.5; DB 23; Length 4115;  
 Best Local Similarity 52.2%; Pred. No. 0;  
 Matches 1555; Conservative 425; Mismatches 802; Indels 197; Gaps 43;



QY 41 GSRGSSGACAMGMLCPSSASNLMLYTRCDAITELTG---HLVPHDGLRWCPES 97  
Db 193 GSRKRVGTACCCBELMWTSGKGG---BRATTSGLERAPRNCILPAGSG----- 229  
QY 98 EAHIPLPRAPEGCPWCRLLIGIGCHLSPOGKLTLPBEPHCLKAPRLRCOS---CK-LAQA 153  
Db 240 -----SGPELDSAPRTARPAPOS-APRESRTAPBEPAP---KRNRSRLGFCRFLPQR 289  
QY 154 PG-----LRAGERSPEBSLGGRRKNVNTAPQOPSPQATVYEPNQPAGTVAISLRAI 206  
Db 290 PGRPRPCLPAPRPAKKTTSANRAAFRPAARHPQPPQYNQTLVPENEAAGTVLRVAQ 349  
QY 207 DPDEGEAGLEYTMDALFDSRSNOFSLDPVTGAVTTABELDRBETKSTHYFRVTAODHM 266  
Db 350 DPDAAGEAGRLVYSIALMNSRSLFSLIDBQSGLRRAALDRBEMRHLYRTAODHS 409  
QY 267 PRSSALATLTLLTDTNDHPVPEOQRYKESLBLEVEGYEVLTVRATDGDAPNNAILY 326  
Db 410 PRLSATTMVAVTVADRNDHSPVEQAOYRETLRENEBEGYPIQLRATDGDAPNNAILY 469  
QY 327 RL- -EGSGSPSEVEPIDRSQVIRTRGVDRBEEVSQOLVTEASDQGRDPGRSTTAA 384  
Db 470 RYVGPAPARAAAFAEIDPRSGLISTSKVDRBHEMSEYELVVEASDQGRDPGRATYR 529  
QY 385 VFLSVEDNDNAPQPSBKRYYVQVREDVTEGAPVLRTASDRDKGSNAVHVSIMSGNAR 444  
Db 530 VHTITVDENDNAPQPSBKRYYVQVREDVTEGAPVLRTASDRDKGSNAVHVSIMSGNAR 589  
QY 445 GQYVLDAGQALDVSPLDYETTEKTYLRVADODGRRPPLSNVSGLVTVQVLDINDAPI 504  
Db 590 GHRFAISLTGELGVVAPLDFEAREEYALIRIADAGRRPPLSNNTGLASIQVVDINHPI 649  
QY 505 FVSTPQATVLSVPLGYLVLAHYQALDADAGNARLEVRLAGHDPPTTINNGCTMIV 564  
Db 650 FVSTPQATVLSVPLGYLVLAHYQALDADAGNARLEVRLAGHDPPTTINNGCTMIV 709  
QY 565 AAELEDEEVDYFSGVEARDHGTPTALTASASVTVLDVNDNPTTOBEPYTRLNEDAA 624  
Db 710 SGLDRESEVHYFVGVEARDHGTPTALTASASVTVLDVNDNPTTOBEPYTRLNEDAA 769  
QY 625 VGTSVTVASANDBASVITYQTSGNTNRBSITQSOGGGLVSLALPLDYKEROYVLA 684  
Db 770 VGTSVTVASANDBASVITYQTSGNTNRBSITQSOGGGLVSLALPLDYKEROYVLA 829  
QY 685 VTASDGTRODTAQIVNVVDANHRPVFOSHTTVVNDPRAGTIVLISANDERTYEN 744  
Db 830 LTHSDRALHDCYVHNITDANTHRPVFOSHTTVVNDPRAGTIVLISANDERTYEN 889  
QY 745 ARTTVEEDSIPOFRIDADTGAATTOAELEDEDOVSYTLATARDNGIFQKSDTYLEIL 804  
Db 890 ARTTVEEDSIPOFRIDADTGAATTOAELEDEDOVSYTLATARDNGIFQKSDTYLEIL 949  
QY 805 VNDVNDNAPQFLDSTYQSGYTEDVPPTVSYLOISATDRDSGLNGRVFTYQSGDGDGDF 864  
Db 950 VNDVNDNAPQFLDSTYQSGYTEDVPPTVSYLOISATDRDSGLNGRVFTYQSGDGDGDF 1009  
QY 865 IVESTSGIRTLRLDRBNVAQVLAAYAVDKGMPARTPMETVTVLDVNDNPTTPEBD 924  
Db 1010 TLEPTSGIRTVRLDRBNVAQVLAAYAVDKGMPARTPMETVTVLDVNDNPTTPEBD 1069  
QY 925 EFDVFEVENSPIGLAVARVATPDDEGTNAQIMYQIVEGNIPREPOLDFSGELTALVUL 984  
Db 1070 EFDVFEVENSPIGLAVARVATPDDEGTNAQIMYQIVEGNIPREPOLDFSGELTALVUL 1129  
QY 985 DYEDREYLVIOATISAPLVSRATVAVRLDRNDNPPVLANFEILFNNTVYNNSSFFPG 1044  
Db 1130 DYEDREYLVIOATISAPLVSRATVAVRLDRNDNPPVLANFEILFNNTVYNNSSFFPG 1189  
QY 1045 AIGRVAHDPDLSLSTYSPERGENLSIVYLNNASTGELKLSRALDNNRFLALIMSTLYSD 1104  
Db 1190 IIGRIYAIYDPDVSDDLPTSPERGENLSIVYLNNASTGELKLSRALDNNRFLALIMSTLYSD 1249

QY 1105 GYHSAQAQALRVTTITTDMLTHSITLRLEDMSPERFLSPILGLFIOAVAATLATPBDHY 1164  
Db 1250 GLHSAVTAQCLVAVVITTEELANSILTVRLNNMOERFLSPILGRFLBGAVALATAEV 1309  
QY 1165 VVFNVRDTPADQGHILANVSISVQPPGPGGP- -PLPESDLOERLYNRSILTAISAQ 1222  
Db 1310 FIFNIQNDTIV- -GTVLANSFSGALAPRGAGAGAGWFSSEBELQOLYRRALARSIL 1368  
QY 1223 RVLPPDNI CLABPCENVRVSVLRFDSAPFLSSSLPPIPHVGLRRCRCPGFG 1282  
Db 1369 DVLPEDNVCILBPCENVMKCVSLRFDSAPFLSSSLPPIPHVGLRRCRCPGFG 1428  
QY 1283 DYCETEVDLCYSRPGCPHRCRSREGYCTLCRDGYTGHCNVSARSAGCTGVCAQGT 1342  
Db 1429 DYCETEVDLCYSRPGCPHRCRSREGYCTLCRDGYTGHCNVSARSAGCTGVCAQGT 1488  
QY 1343 CYNLVGFKCDPCSGD- -PEKPYCOVTRSPHSPFTFRGLRQRFHTLALSPATKED 1401  
Db 1489 CYNLVGFKCDPCSGD- -PEKPYCOVTRSPHSPFTFRGLRQRFHTLALSPATKED 1548  
QY 1402 GLLVYNGRNEKHDPVLEVLQEOYQVTFPSAGESTTVSPFVPGVSDQOMHTQKYN 1461  
Db 1549 GLLVYNGRNEKHDPVLEVLQEOYQVTFPSAGESTTVSPFVPGVSDQOMHTQKYN 1608  
QY 1462 KPLAQGTGLPQSPSEQKVAVVTVDCDITGVALLPFSVLGNYSCAAQGTGSKSLDLTG 1521  
Db 1609 KPLAQGTGLPQSPSEQKVAVVTVDCDITGVALLPFSVLGNYSCAAQGTGSKSLDLTG 1668  
QY 1522 PLLGQVPLPESFPVRMOPVGCNRENTQVDSRHITMDAPFANNGTVPCCPAKXVCDN 1581  
Db 1669 PLLGQVPLPESFPVRMOPVGCNRENTQVDSRHITMDAPFANNGTVPCCPAKXVCDN 1728  
QY 1582 TCHNGSTCVNOMDASCPGELFGSKCAQEMANQHTLSSLVMM- -GLSLPIQOPWL 1640  
Db 1729 TCHNGSTCVNOMDASCPGELFGSKCAQEMANQHTLSSLVMM- -GLSLPIQOPWL 1788  
QY 1641 SLMPRTQADGVLTQAITRGRSTITLQIREGHVLSV- -EGTLQASSILREGRANDGDW 1699  
Db 1789 SLMPRTQADGVLTQAITRGRSTITLQIREGHVLSV- -EGTLQASSILREGRANDGDW 1847  
QY 1700 HHAQAL- -GASGFGHAIL- -SPDYGOORAEAGNIGPLHGLHSNTYVGI- -PGPAGV 1754  
Db 1848 HHAQAL- -GASGFGHAIL- -SPDYGOORAEAGNIGPLHGLHSNTYVGI- -PGPAGV 1907  
QY 1755 ARGFGCQGVAVSPTRPGVNS- -LDPSHGBSLNVGQSLDPPCPSNCPNAYSQSNMD 1813  
Db 1908 ARGFGCQGVAVSPTRPGVNS- -LDPSHGBSLNVGQSLDPPCPSNCPNAYSQSNMD 1965  
QY 1814 SYSQSCDGGYGDNCTNVCIDLPCEHOSVCTRKPSAPRGYTCBPNTLGPYCETRIDP 1873  
Db 1966 SYSQSCDGGYGDNCTNVCIDLPCEHOSVCTRKPSAPRGYTCBPNTLGPYCETRIDP 2025  
QY 1874 CPGWGHPTGCPKCDVSKGDPDPCNTSGECHKENHTRPQSPPTCLLDCYTGSL 1933  
Db 2026 CPGWGHPTGCPKCDVSKGDPDPCNTSGECHKENHTRPQSPPTCLLDCYTGSL 2085  
QY 1934 RYCDPEBQCPKCPVITROCDRCNPRFAYTTNCCFNYSCPAIALGIMPTTRGCL 1993  
Db 2086 RYCDPEBQCPKCPVITROCDRCNPRFAYTTNCCFNYSCPAIALGIMPTTRGCL 2145  
QY 1994 PAAPCPKGSFGTAVRCDERHGMPLPNTLFCNTSTTFBELGAFARLORNSGLDSCGSO 2053  
Db 2146 PAAPCPKGSFGTAVRCDERHGMPLPNTLFCNTSTTFBELGAFARLORNSGLDSCGSO 2205  
QY 2054 QIALILRNATOTHTAGYFSGDYKAVAYQALTRLLAHSTORGGLSATOVHFTENLRVGS 2113  
Db 2206 QIALILRNATOTHTAGYFSGDYKAVAYQALTRLLAHSTORGGLSATOVHFTENLRVGS 2265  
QY 2114 ALLDTPANKHWEIL- -QOTEG- -GRAMLOHYEYASALAOHMTYVSPFTVYVNTVI 2169  
Db 2266 ALLDTPANKHWEIL- -QOTEG- -GRAMLOHYEYASALAOHMTYVSPFTVYVNTVI 2325  
QY 2170 SVVRLDKGNF- -AGAKLPRVYA- -LRGQPPDLFTTVILPESVPERTPPVVAPAGGEAQ 2225

```

Db      2326 SIDRMHPSPPRGARPRYHSNLPFGQAMDPHTVLLPSQSPRSPSEVLPTSSSIN 2385
Qy      2226 EPEELARRGRHRELSQGEAVAVIYRTAGLIPNNYBPDKSLAVPKRPINTPVSI 2285
Db      2386 STTSVPPPPAPEPEPGISIIILLVYRTGLLPQAFQEREGALPQNPVNSIPVSV 2445
Qy      2286 SVHDEELPRALDKPVTVQFRLLETEERTKPCVPMNSILVSGSGMSARCEVVFNN 2345
Db      2446 AVFHGNFPLRGLESPISEFLRLQTNNSKALCYQMDPGLACQGWTAHDCELVHN 2505
Qy      2346 ESHVSCQCNHMTSPAVLMDVSRRE--NGEILPKLTYVALGVTLALITPEPFLTRI 2403
Db      2506 GSHARCRCRTGTFGVMDASPREERLELAVFTHVVAVALVTAIILSLRS 2565
Qy      2404 LRNNGHGRNLTALAGLQVFLGINQ-----ADLP 2436
Db      2566 LKSNVGHANVANVAALGVALLFLGLGHRTHNOVQDQGTCLVMTLLAQEWGNSGE 2625
Qy      2437 PACTVAIILHPLYLCTPSWALLBALHLYRALTEVDVNTGPMRYMVGMPAFITGL 2496
Db      2626 LVCTAAVAILHLYFTSTFAMLFGVGLHLYMVEPRINVRGNRFTHALGMPAVLLEL 2685
Qy      2497 AVGLDEPGIGNPFCMLSTYDTLINSRPAFVAVASVFLYLAAPASCAQROGPEKK 2556
Db      2686 AVGLDEPGYGNPFCWISVREPLIWSFAGFVVLVIVNGMFTLAARTSCSTQREAKKT 2745
Qy      2557 GPVSGLGQSPAVLILATWLTLLALSVNSPTLLFHTLPATNCIQGPFIPLSVVLSKEV 2616
Db      2746 SALRTKRSFLILVLSASLFLGLVLANHSLILAFHLLAGLCGLQGLAVLLPLCVLNADA 2805
Qy      2617 RKALKIAC--SRKSPDPALTTKSTLTSSVNCSPYVADGRLYQ--PYGDSAGSLHSTRSGK 2674
Db      2806 RAAWMPACICRKAAPREARPAELGPAVNNALPESGLIRITLASITVSSVSAISR 2865
Qy      2675 SQP-----SYT--PFLARESA-----LNPQGPFGIG-----DPGSLFLEGQO 2712
Db      2866 TDQDSQQRGASLYRDVNLVHNGSAADHTDHSLOAHGAPDLDVAMFHRDGA----- 2917
Qy      2713 QHDPDTSDSLLEDQSGSYASTHSSDSSEEESEEEBAAPRGQWDSLILGQABRL 2772
Db      2918 ----DSDSDSLSEERSLSTSSSEEDNGRTGRGFORPLCAAAQ-----SERL 2963
Qy      2773 PLHSTPKDGGPGPKAPWG-----DFGT-----TAKSSGNGAPEERLRE 2813
Db      2964 LTH--FKDYDGMNLLSYWPAIGCEBAPCALQWGERRLGLTSDAANNQPDAL-T 3020
Qy      2814 NGDALSRBGLGDLPGSSAOPHKGILKKCLPTISEKSLIRLP-----LEQCTG 2863
Db      3021 SGD-----ETSL-----GRAQRQRKGIKNR-----LOYPLVQTRGAPELSWCRA 3061
Qy      2864 SSRGSA-----SGSRGGPPPPPPROSLQELN 2893
Db      3062 ATIGHRAVPAASGYRTAGGCTGSLQPARVYSRRQLD 3100

```

```

KW      glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
KW      osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
KW      infection; cell proliferative disorder; actinic keratosis; myeloma;
KW      arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotoxic;
KW      neutrotective; dermatological.
OS      Homo sapiens.
FH      Key
FH      Peptide
FT      1..31
FT      /label= Signal_peptide
FT      32..3298
FT      /note= "Mature human extracellular matrix and cell
FT      adhesion molecule (XMAP)"
FT      329..423
FT      /note= "Cadherin domain"
FT      437..535
FT      /note= "Cadherin domain"
FT      549..641
FT      /note= "Cadherin domain"
FT      655..746
FT      /note= "Cadherin domain"
FT      760..848
FT      /note= "Cadherin domain"
FT      862..952
FT      /note= "Cadherin domain"
FT      965..1057
FT      /note= "Cadherin domain"
FT      1071..1159
FT      /note= "Cadherin domain"
FT      1178..1265
FT      /note= "Cadherin domain"
FT      1378..1431
FT      /note= "EGF domain"
FT      1438..1469
FT      /note= "EGF domain"
FT      1478..1512
FT      /note= "EGF domain"
FT      1536..1553
FT      /note= "EGF domain"
FT      1542..1704
FT      /note= "Transmembrane domain"
FT      1725..1756
FT      /note= "Laminin G domain"
FT      1792..1920
FT      /note= "Laminin G domain"
FT      1931..1962
FT      /note= "EGF domain"
FT      1966..2000
FT      /note= "EGF domain"
FT      2110..2137
FT      /note= "GPCR signature"
FT      2525..2545
FT      /note= "GPCR signature"
FT      2588..2613
FT      /note= "Transmembrane domain"
FT      2659..2689
FT      /note= "GPCR signature"
FT      2712..2730
FT      /note= "Transmembrane domain"
FT      2741..2761
FT      /note= "Transmembrane domain"
PD      WO200142285-A2.
PN      14-JUN-2001.
XX      Human extracellular matrix and cell adhesion molecule-21 (XMAP-21).
XX      Human: extracellular matrix and cell adhesion molecule; XMAP;
XX      gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
XX      Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
XX      sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
XX      inflammatory disorder; acquired immune deficiency syndrome; AIDS;
XX      Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
XX      Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
PA      (INCY-) INCYTE GENOMICS INC.

```



QY 1462 KPLIGTGTLPOGSEBQKAVAVTTVDGCDTVALRFGSVLGNYSCAAQGTGSGSKSLDLTG 1521  
 DB 1609 KPRFDALGAGQSPKDKAVAVLSTVDDCDVAVALFGABIGNYSCAAAGVQTSKSKSLDLTG 1668  
 QY 1522 PLLIGGPDLPBESPVMROFVCGKRLQUDSRIDMADPLANNGTTPGCPAKKNVCDN 1581  
 DB 1669 PLLIGSVPNLPENPFVHKDPTGCMRDIAHIDGRVDMAAVANNGTAAQAKLHPCDSG 1728  
 QY 1582 TCHNGTICVNMWDAPSCCEPLFGSGKSCAQMANNPOHFLGSLVAMH-GLSLPIQPMYL 1640  
 DB 1729 PCNRSGCSERMSGFSDCPFGGKDCQLTMALPHFRNGITLSMFGSDMAVSVMYL 1788  
 QY 1641 SLMFRTQADGVLQALITRGSTITLQLRBGVNL-SV-EGTGLQASSLRLEPGRANDGW 1699  
 DB 1789 GLAFRTATQGVLMQVQAPHSITLLCOLDRGLSVLTRGSG-RASHLLDQVTVSDGRV 1847  
 QY 1700 HMAQLAL--GASGGRGHAL--SPDYQOQABEGMLGRLHGLHLSNTTGGT-PGAPGCV 1754  
 DB 1848 HDLRLLEQEBGGRGHHVLMVSLDFSLFDOTMAVSELOGLKVKQLHVGGLPGSABEA 1907  
 QY 1755 ARGFRGCTQGVAVSDTPEGVNSLDPHSGESINVEQGSLLPDCSNPCPANSYCSNDWDS 1814  
 DB 1908 PGLVGCIG-----PPS-----ECGPGCVTNACASGFCPRHADCRLMQT 1948  
 QY 1815 YSGSCDPGYIGDNCTVCDLNPCEHSGVCTRKPSAPHGTYCECPNYLGYCETRLDQPC 1874  
 DB 1949 FSCCTCGGYGPGCVDACLNPPCONQSCRLPQAPHYTCDGCVGFHHCERHMOQC 2008  
 QY 1875 PRGMWGHPTGPGCNDVSKGPRDPCNTSGECGCKENHYRPGSPFTLLDCDCTPGSLR 1934  
 DB 2009 PRGMWSPCTGPGCNDVHKGDPCNKTNGQCKEKEHYRPRGSDCLPCDCTPVGSTR 2068  
 QY 1935 VCDPEBQCPCKGVLIGQDCRCDNPPAEVYTNCEVYNVSCPRALBAGIWMPTREGLP 1994  
 DB 2069 SCAPHSGQCPRGALGRQCNCSOSPFAVYTAASCRLYACCKSLASGVMWQTKRGVL 2128  
 QY 1995 AAAPCPKGSF-----GTAVHCDEHKGMLPPNLPNCTSTPFSBLKFAERLQNNESGLDS 2049  
 DB 2129 ATVPCEPRGALGRGAGAVRLCDEAOGMLBPDLPNCTSPAFRELSLLDGLINKTALDT 2188  
 QY 2050 GRSQOLALLRNATOPHTAGYFGSDVVKAYAYOLATRLAHSTORGFGLSATODVHTENIL 2109  
 DB 2189 MEAKKLAQRLEBYTGHTDHYPSQDVRYTARLLAHLALPESHQGFGLTATODAHFENIL 2248  
 QY 2110 RVGSALLDTRANKHWEIL-QOTBG--GTAMLQHYEAVASALAQNMRHTYLSPTIYTP 2165  
 DB 2249 WAGSALLAPETGDMALGQAPGSGSAGLVRLHEVAATLARNNELTYLNPGLVTP 2308  
 QY 2166 NIVISVRLDKNF--AGAKLPREYA--LAGEOPDLETTYILPESVRETPEVVRPAPG 2221  
 DB 2309 NIMLSIDRMHPPSPRGARRYPRYHNLFRGODAMPHTVLLPSQSPRPSSEVLPTSS 2368  
 QY 2222 GBAOPEELARORRHEPLOSGBAVASVITYRTLAGLLPHNYDPODKSLVPPKPIINTP 2281  
 DB 2369 SIENSTSSVPPPAPEPPEPGISITILLVYRILGILLPQFOAERRGARLPQNPVWNSP 2428  
 QY 2282 VASISVHDEELLPRALDKPVTVQFLLLEETERTKPICVFNHNSILVSGTGWMSARGCEV 2341  
 DB 2429 VVSVAVHGNPLRGILLESPLISLEFRLQYANRSKALCYQMDPPGLAEQGVWTAARDCEL 2488  
 QY 2342 VPRNESHVSCQCHMHSFALIMVSRRE--NGBILPKLITYVALGYTLAALLITPEFLT 2399  
 DB 2489 VHRNGSHARCRCRTGTFFGVLMDSPPERLEGDELLAVFTHVAVASVALVTLTAAILL 2548  
 QY 2400 LRLILSNQHGIRRLTAALGALQVLGLGINOALPEFACTVAILLHFLYLCTPSWALL 2459  
 DB 2549 SLSKSXNVAGIHANVAALGVABELFLGLGIRHYNQVCTAVAILLHYFFLSTPAMLY 2608  
 QY 2460 BALHLTALTEVADVNTGPKRFTYMLGMPAFITGLAVGLDBEGYGNPDCWLSIYDTL 2519  
 DB 2609 OGHLTYMOVPRVNDGAMRFYHALGMPAVALLGLAVGLDBEGYGNPDCWISVHEPL 2668  
 QY 2520 IWSRPAQVAVAVSNVFLYLILAAASCA-AQROGFEKKGPVSGLOPSPAVILLLSATWLL 2578

DB 2669 IWSRPAQVAVAVSNVFLYLILAAASCA-AQROGFEKKGPVSGLOPSPAVILLLSATWLL 2578  
 QY 2579 ALLSVNSDTLHRYLFPATCNICQGPFTILSYVLLASVRYKALKALAC-SKPSPDPLTTK 2637  
 DB 2727 GLAAVHSHLAFHYLALGCGIGLAVLLHPCVLNDAANAAMWPACTIGRKAAPREBARPAV 2786  
 QY 2638 STLSYNGSPYADGRLYO-PYGDAGSLHSTRSGKSP-----SYI--PFLREE 2687  
 DB 2787 GLAPGAYNTALPESBGLIRITLGASTVSSVSSAGSGRTQDDSDQGRGYLKNVLYRG 2846  
 QY 2688 SA-----LMPGQGPGLG-----DPGSLPLBGDDQHDPTDSDSLSEDDGSGY 2734  
 DB 2847 SAADHTDHSLOAHAGPTDLVAMFHRDAGA-----DSDESDLSLEBERSLSI 2894  
 QY 2735 ASTHSDSEBEBEBEBEBEAAFGEOGMDSLPARGARLPLASTPRDGGGPGKAPWPG-- 2792  
 DB 2895 PSSESDNGRTGRVQRPICRAQ-----SERLLTH--PKVDGNDLLSYVPALG 2942  
 QY 2793 -----DPGT-----TAKESSGNGAPEERLENGDALSRGSLGPLPGSSAOPH 2835  
 DB 2943 ECEAAPCALQTWGSEBRRLGDLTSKQANNNQDPAL-TSGD-----FTSL-----GRAQROR 2993  
 QY 2836 KGLKKKCLPTISEKSLRLP-----LEOCTGSSRGSSA-----SEGSRG 2876  
 DB 2994 KGLKKR-----LQYPLVPTQTRGAPELSWCBAATLGHRAVPAASVGRIVAGGT 3042  
 QY 2877 GPPPPPPRQSLQEOULN 2893  
 DB 3043 GSLSDPASRYSSRBOLD 3059

Search completed: February 11, 2004, 15:54:56  
 Job time : 74 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 11, 2004, 15:53:34 ; Search time 31 Seconds  
(without alignments)  
3989.504 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545

Sequence: 1 MRSPTVGVLPPTPPPLLL.....AGTVDSDSGSRFLFFNTLH 2923

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1004.5	6.5	884	2	US-08-465-976A-2
2	1004.5	6.5	884	2	US-08-982-412-2
3	846	5.4	1026	1	US-07-998-003A-95
4	846	5.4	1026	1	US-08-453-274B-95
5	846	5.4	1026	1	US-08-453-695A-95
6	846	5.4	1026	1	US-08-453-161A-95
7	846	5.4	1026	2	US-08-453-702A-95
8	846	5.4	1026	3	US-09-099-639-95
9	846	5.4	1026	5	PCT-US93-12588-95
10	846	5.4	1026	5	PCT-US93-08071-95
11	846	5.4	1203	1	US-07-998-003A-103
12	846	5.4	1203	1	US-08-453-274B-103
13	846	5.4	1203	1	US-08-453-695A-103
14	846	5.4	1203	1	US-08-268-161A-103
15	846	5.4	1203	2	US-08-453-702A-103
16	846	5.4	1203	3	US-09-099-639-103
17	846	5.4	1203	5	PCT-US93-12588-103
18	846	5.4	1203	5	PCT-US93-08071-103
19	735.5	4.7	1466	4	US-09-262-537-58
20	727	4.7	1466	4	US-09-262-537-20
21	719	4.6	1403	4	US-09-262-537-6
22	711.5	4.6	780	4	US-09-262-537-10
23	711.5	4.6	1114	4	US-09-262-537-34
24	711.5	4.6	1177	4	US-09-262-537-2
25	691	4.4	1123	4	US-09-262-537-4
26	628.5	4.0	797	1	US-08-453-695A-112
27	628.5	4.0	797	1	US-08-268-161A-112

28	628.5	4.0	797	2	US-08-453-702A-112	Sequence 112, App
29	628.5	4.0	797	3	US-09-099-639-112	Sequence 112, App
30	628.5	4.0	797	5	PCT-US93-08071-112	Sequence 112, App
31	624.5	4.0	787	1	US-08-453-695A-110	Sequence 110, App
32	624.5	4.0	787	1	US-08-268-161A-110	Sequence 110, App
33	624.5	4.0	787	2	US-08-453-702A-110	Sequence 110, App
34	624.5	4.0	787	3	US-09-099-639-110	Sequence 110, App
35	624.5	4.0	787	5	PCT-US93-08071-110	Sequence 110, App
36	611.5	3.9	836	1	US-07-998-003A-105	Sequence 105, App
37	611.5	3.9	836	1	US-08-453-274B-105	Sequence 105, App
38	611.5	3.9	836	1	US-08-453-695A-105	Sequence 105, App
39	611.5	3.9	836	1	US-08-268-161A-105	Sequence 105, App
40	611.5	3.9	836	2	US-08-453-702A-105	Sequence 105, App
41	611.5	3.9	836	3	US-09-099-639-105	Sequence 105, App
42	611.5	3.9	836	5	PCT-US93-12588-105	Sequence 105, App
43	611.5	3.9	836	5	PCT-US93-08071-105	Sequence 105, App
44	611.5	3.9	904	1	US-07-998-003A-97	Sequence 97, App1
45	611.5	3.9	904	1	US-08-453-274B-97	Sequence 97, App1

## ALIGNMENTS

RESULT 1  
US-08-465-976A-2  
; Sequence 2, Application US/08465976A  
; Patent No. 5869632  
; GENERAL INFORMATION:  
; APPLICANT: SOPPET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: RUBEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: CABELLA, BYRNE, BAIN GILFILLAN, CECCHI  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,976A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY F  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-444  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 994-1700  
; TELEFAX: (201) 994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 884 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-465-976A-2

Query Match 6.5%; Score 1004.5; DB 2; Length 884;  
Best Local Similarity 37.3%; Pred. No. 3e-63;  
Matches 254; Conservative 103; Mismatches 213; Indels 111; Gaps 20;  
CY 2278 INTVSISVHDBELLPRALDKVTVQFPLREERTKPLCVPMNISIVSGSGGNSAR 2337  
DB 1 MRSPTVGVLPPTPPPLLL.....AGTVDSDSGSRFLFFNTLH 2923

```

QY 2338 GCEVFRNESHVSCQCNHMTSPFVLMVDSRRE--NGEILPLKTLTYVALGVTALALLTF 2395
DB 61 DCELVHRNGSHARCRCSTRTGTFGVLMDSRPRELEGBLELAVFTHVVAVALVLT 120
QY 2396 FFLTLRLRSNOHGRIRNLTALGLAQLVFLGINQADLPFACVIAIILHFLYLCFS 2455
DB 121 AILLSRLSKSVNRGIHANVAALGVABELFLGIHRTNQVCTAVAILHFLSTPA 180
QY 2456 WALLBALHYALTEVRDVTNGPMPRTYMLGMGVPAFTGLAVGLDPGCGNDFCWTISI 2515
DB 181 WLFVQGLHYRQVVEPRVDNGRMRFYHALGMGVPAFTGLAVGLDPGCGNDFCWTISV 240
QY 2516 YDTLWSPAGPVAFVAVNSVFLYIILABASCA--AQROGFEKKGPVSGLOPSPAVILLISA 2574
DB 241 HEPLWSPAGPVVLYIVNMGTMFLIARISCTGQREA--KTSALTLRSSFLILLVISA 298
QY 2575 TWLALLSVNSDTLFLHYLPATCNCIQGPFTFLSYVLSKEYRKALKLAC--SRKSPDPA 2633
DB 299 SWLFGILAVNHSLAFHYLHAGLCGLAVILLFCVLANDAARAAMPACIGRKAAPERA 358
QY 2634 LTTKSTLTSYNCPSPYADGRLYQ--PYGDSAGSLHSTSRGSKOP-----SYL--PFL 2683
DB 359 RPAPGLGPGAYNNTALFEBSGLIRITLGAISTVSVSSASGRYQDDSQGRGSYLRDNVYL 418
QY 2684 LREESA-----LNPQGGPPGLG-----DPGSLFLEGGDDQDHDPTDSDSLSLBEDQ 2730
DB 419 VHRGSAADHTDHSLOAHAGPTDLDVDMFHRDAGA-----DSDSDSLSLSEER 466
QY 2731 SGSYASTHSDSEEBEERBEERAAFPBQGMDSLLGPAERLPLHSTPKDGGPGGKAPW 2790
DB 467 SLSTPSESEDNRTGRFGRPLCRAGQ-----SERLITLH--PKVDGNDLLSLTW 514
QY 2791 PG-----DGT-----TAKESGNGAPBERLRBNGDALSRGSLGFLPGSS 2831
DB 515 PALGCEAAPCALQTWGSEERLGLDTSKDAANNQDPDAL--TSGD-----ETSL--GRA 565
QY 2832 ACPHKKILKKCLPTISEKSLRLP-----LEOCTGSSRGSSA-----SE 2872
DB 566 QRRKGIILKNR-----LQYPLVPQTRGAPELSMCRATLGHRAVPAASYGRITYA 614
QY 2873 GSRGPPPPPPROSLQROLN 2893
DB 615 GCGTSLSQPASRYSSREOLD 635

RESULT 2
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE,
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982.412
; FILING DATE:
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PFI181PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-982-412-2

Query Match 6.5%; Score 1004.5; DB 2; Length 884;
Best Local Similarity 37.3%; Pred. No. 3e-63;
Matches 254; Conservative 103; Mismatches 213; Indels 111; Gaps 20;

QY 2278 INTPVVISYHDEBELPRALDKPVTVQPRLLTEERTKPICVFNHSLIVSGTGWGAR 2337
DB 1 MNSPVSAVAVPHGRNFTLGILSPISLEFRLLQTNRSKALCVQMDPGLBQHGVMYAR 60
QY 2338 GCEVFRNESHVSCQCNHMTSPFVLMVDSRRE--NGEILPLKTLTYVALGVTALALLTF 2395
DB 61 DCELVHRNGSHARCRCSTRTGTFGVLMDSRPRELEGBLELAVFTHVVAVALVLT 120
QY 2396 FFLTLRLRSNOHGRIRNLTALGLAQLVFLGINQADLPFACVIAIILHFLYLCFS 2455
DB 121 AILLSRLSKSVNRGIHANVAALGVABELFLGIHRTNQVCTAVAILHFLSTPA 180
QY 2456 WALLBALHYALTEVRDVTNGPMPRTYMLGMGVPAFTGLAVGLDPGCGNDFCWTISI 2515
DB 181 WLFVQGLHYRQVVEPRVDNGRMRFYHALGMGVPAFTGLAVGLDPGCGNDFCWTISV 240
QY 2516 YDTLWSPAGPVAFVAVNSVFLYIILABASCA--AQROGFEKKGPVSGLOPSPAVILLISA 2574
DB 241 HEPLWSPAGPVVLYIVNMGTMFLIARISCTGQREA--KTSALTLRSSFLILLVISA 298
QY 2575 TWLALLSVNSDTLFLHYLPATCNCIQGPFTFLSYVLSKEYRKALKLAC--SRKSPDPA 2633
DB 299 SWLFGILAVNHSLAFHYLHAGLCGLAVILLFCVLANDAARAAMPACIGRKAAPERA 358
QY 2634 LTTKSTLTSYNCPSPYADGRLYQ--PYGDSAGSLHSTSRGSKOP-----SYL--PFL 2683
DB 359 RPAPGLGPGAYNNTALFEBSGLIRITLGAISTVSVSSASGRYQDDSQGRGSYLRDNVYL 418
QY 2684 LREESA-----LNPQGGPPGLG-----DPGSLFLEGGDDQDHDPTDSDSLSLBEDQ 2730
DB 419 VHRGSAADHTDHSLOAHAGPTDLDVDMFHRDAGA-----DSDSDSLSLSEER 466
QY 2731 SGSYASTHSDSEEBEERBEERAAFPBQGMDSLLGPAERLPLHSTPKDGGPGGKAPW 2790
DB 467 SLSTPSESEDNRTGRFGRPLCRAGQ-----SERLITLH--PKVDGNDLLSLTW 514
QY 2791 PG-----DGT-----TAKESGNGAPBERLRBNGDALSRGSLGFLPGSS 2831
DB 515 PALGCEAAPCALQTWGSEERLGLDTSKDAANNQDPDAL--TSGD-----ETSL--GRA 565
QY 2832 ACPHKKILKKCLPTISEKSLRLP-----LEOCTGSSRGSSA-----SE 2872
DB 566 QRRKGIILKNR-----LQYPLVPQTRGAPELSMCRATLGHRAVPAASYGRITYA 614
QY 2873 GSRGPPPPPPROSLQROLN 2893
DB 615 GCGTSLSQPASRYSSREOLD 635

RESULT 3
US-07-998-003A-95
; Sequence 95, Application US/07998003A
; Patent No. 5643781
; GENERAL INFORMATION:

```



APPLICANT: Suzuki, Shintaro  
 TITLE OF INVENTION: Protocadherin Materials and Methods  
 NUMBER OF SEQUENCES: 107  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ADDRESSER: Bicknell  
 STREET: 20 South Clark Street  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/998,003A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 564378land, Greta E.  
 REGISTRATION NUMBER: 35,302  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/346-5750  
 TELEFAX: 312/984-9740  
 TELETYPE: 25-3856  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-998-003A-95

Query Match 5.4%; Score 846; DB 1; Length 1026;  
 Best Local Similarity 29.1%; Pred. No. 1e-51;  
 Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPPNANI-----LYRLBSGGSPSEVFIEDPRSG-VI 349  
 DB 33 GHATRVVYKVPBEQPPNTLIGSLADYGPDPVGHLYKLEVG-----APYLAVDKGTGIF 87  
 QY 350 RTRGVDREEVESYQ-----LTVASDQGRDPGPRSTTAAVFLSYEDDNDNAPO 398  
 DB 88 TTESIDREGLRECONQLFGDPCLIEFEVSTIDLVQNASPRLLEQOI--EVQDINDNTFN 145  
 QY 399 FSEKRYVQVREDVTPGAFVLRVTASDRDKGSNAVHYSIMSGNARGQFYLDAGTALDV 458  
 DB 146 PASPVITLAIPEMTNIGSLFPIPLASDRDAGNVASVELQVAEDQ-----BEKPOLIV 200  
 QY 459 VSPLYETTKETTLKVRADGGRPPISNVSLVTVQVLDINDNAPIFVSTPPQATVLSV 518  
 DB 201 MGNLDRERWDSDTLTKVQDGSPPPA--TSALLRVTVDLTNNAPKFERPSEYAELESENS 259  
 QY 519 PLGYLVLAHQALDADAGNARLEVLAVGHPF-----FTINNGTGMISVAELDERE 572  
 DB 260 PLGHSVIOVANDSDQGANARLEIYTF---HQAPEVRRLLRLDRNTGLITVQGVNDESD 315  
 QY 573 VDFVSGVEARDHGTPALTFASASVSTVLDVNDNPTF-----TOPEYTVRLNEDAA 624  
 DB 316 LSTLRFPSVLAKDRGNPKSARAAQVVTVDMDNAPFTIERIGLVTHQDGMANISEDA 375  
 QY 625 VGTSTVTVTSAYVRD--AHSVITYQITSGTNRFPSTTSGSGG-----LVSLALPLDYKL 677  
 DB 316 ERTAAALVQVSDRDEBENAAVTC--VVAQVPPQLQASSTGDSKXKVFLLQTTTPIIDYEX 434  
 QY 678 BRQYVLAATASDGR---QDTAQIVNVNTDANTHRVPQSSHYTVNVNEDRAGTIVVLI 734  
 DB 435 VQDYITIEIYAVDSGNPPLASTNSLKVQVVDNDNAPVFQSTVEVAFENKPGEVIAEI 494  
 QY 735 SATDEDTGENARITYFMEDSIPO-----FRIDADGAVTTQAEILDYEDQVSYTLAITAR 788

DB 495 TASHADSSNMEIYYSL-----PEPAKGLFTISPETGSIQVKTSLDRQRSEYELKAVAA 551  
 QY 789 DNGIPQSDTYTYLILVNDVNDNAPQPLRDSYQGSYEDVPEFTSVLOISATDRDSGNG 848  
 DB 552 DRGSPSLQGTATVAVNLVDCNDNDPKFPLSGYNSVNMENMPLSPVGAVTVIIDGKGNA 611  
 QY 849 RVFTYFGDGDGDDGDFIYESTSGIVRTLRLDRNVAGVYLRAVAVDGMPPARTMEVT 908  
 DB 612 QVQLSV---EODNGDPVYQNGTGITLSLSPFRQOQSTTYQLKAVDGVPPRSAYVVT 668  
 QY 909 VTVLVDNDNPEVF--BODEPDPVEENSPIGIAVARVATADPDGRTNAQIMQIVEGNIP 966  
 DB 669 INVLDENDNAPYIAPASNTSHKLLTPQTRIGETVSQVAAEDPDSGVNALLYSIAGNIPY 728  
 QY 967 EVFQDLFSGELTLVLDLDYDRPEYVLYVIOATAPLVSRAVTRVRLDRNDNPPVLCN- 1025  
 DB 729 GLPQIGSHSGAITTEKEIERRHGHRLV-----VKVSDRG-KPPRYGTA 772  
 QY 1026 -FEILFNMYVNRSSSPFGCAIGRPAH-----DPDISDLTY--SPERGENEL 1070  
 DB 773 LVHLVNETLANRK-----LLETLLGHSLDTPVDIDDIADGPEYERSKQGNIL 820

RESULT 4  
 US-08-453-274B-95  
 Sequence 95, Application US/08453274B  
 Patent No. 3663500

GENERAL INFORMATION:  
 APPLICANT: Suzuki, Shintaro  
 TITLE OF INVENTION: Protocadherin Materials and Methods  
 NUMBER OF SEQUENCES: 107  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borum  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/453,274B  
 FILING DATE: 30-MAY-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5663300and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32660  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELETYPE: 25-3856  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-453-274B-95

Query Match 5.4%; Score 846; DB 1; Length 1026;  
 Best Local Similarity 29.1%; Pred. No. 1e-51;  
 Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPPNANI-----LYRLBSGGSPSEVFIEDPRSG-VI 349  
 DB 33 GHATRVVYKVPBEQPPNTLIGSLADYGPDPVGHLYKLEVG-----APYLAVDKGTGIF 87  
 QY 350 RTRGVDREEVESYQ-----LTVASDQGRDPGPRSTTAAVFLSYEDDNDNAPO 398

```

Db      88 TTETSIDREGLRCONQLPDPCILEFEVSIITDLVONASPRLLBQOI--EVQDINDINTPN 145
Qy      399 FSEKRYVVOVREDVTPGAPVLRTVATSDRDKGSNAVVHYSIMSGNARGQFPLDQGTALDY 458
Db      146 FASPVITTLAIPENTNIGSLFPIPLASDRDAGPNVASYELQVAEDQ-----BEKQPOLIV 200
Qy      459 VSPIDVETTKETTLRVRAQDGRPPLSNVGLVTVQVLDINDNAPIFVSTPQATVLESV 518
Db      201 MGNIDREKRWSDYDLTIKVQDGGSPRA--TSALLRVTLVDINDNAPKERSYEAEISENS 259
Qy      519 PLGIVLVIAQALDADGNARLEFRLAGVGHDP-----FTINNGTGMISVAELDREE 572
Db      260 PIGHSVIOVKANDSOGANAEIETFP---HOAEVVRRLRLDNTGLITVQGPVDRED 315
Qy      573 VDFYSGVEARDHGTPALITASASVTVLVDVNDNPTF-----TOPEYTVRLNEDAA 624
Db      316 LSTLRPSVLAKDGTNPKSARAQVTVKMDNNAPIETIRGIGLVTHODGMANISEDA 375
Qy      625 VGTSVTVASAVRD--AHSVITYQITSGNTRNRPISITSGGG---LVSLALPLDYKL 677
Db      376 EETAVALLVQVSDRDEGENAAVTC--VVAGDVPOLRQASETSGDSKKKYFLQTTPLDYEK 434
Qy      678 ERQYVLAVTSADGR---QDTAQIVNVNTDANTHRPVQSHYTVANNEDRPACTTVLI 734
Db      435 VKDYTLIEIVAVDSGNPPLSSTNSLKVQVVDVNDNAPFTQSVTEVAPENNKKGEVIAEI 494
Qy      735 SATDEDTGENARITYFMEDSIPQ-----FRIDADTGAVTTOALEYDQVSYTLAITAR 788
Db      495 TADADSGSNALVYSLE---PEPAKGLFTISPETGEIOVKTSLDRQRESIELKVVA 551
Qy      789 DNGIPQKSDTTLVLELVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLOISATDRDGLNG 848
Db      552 DRGSPSLQGTATVLLVNLDCNDNDPKFMLSQNFVSNEMNPLSPVGMVTVIDGKGENA 611
Qy      849 RIVFYTGQGDGDDGPIVESTSGIVRTLRRLDRENAQVYLRAVADKGMPPARTMEVT 908
Db      612 QVQLSV---EQDNGDFVIOGTGTLISLSLSPRQOSTYTFOLKAVDGVGPPRSAYVGT 668
Qy      909 VTVLDVNDNPPVF--EODEFDVFEENSPIGLAVARTATPDDEGTAQIMYOIVEGNIP 966
Db      669 INVLDENDNAPYITAPNTSHKLLTPQTRIGETVSOVAEDPDSGNALIVISAGNPY 728
Qy      967 EVFQDLIFSGELTALVDLDYEDRPREVYLVIOQTSAPLVSRATVHRLDRNDNPPVLGN 1025
Db      729 GLVQIGSHSGAITLKEKIEIRRHGHLRLV-----VXKSDRG--KPPRYGTA 772
Qy      1026 -FELFNNTVNTSSSPGGAIGRPVH-----DPDISLTY--SFEKGNEL 1070
Db      773 LVHLYVNETIANRT-----LLETGLGSLDTPLDIDIDAGDPEYERKQKGNITL 820

RESULT 5
US-08-453-695A-95
; Sequence 95: Application US/08453695A
; Patent No. 5708143
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; ADDRESSER: Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-95

Query Match      5.4%; Score 846; DB 1; Length 1026;
Best Local Similarity 29.1%; Pred. No. 1e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

Qy      305 GREVLTTRATDGDAPNANI-----LYRLIESSGSGSPSEVFELIDPRSG-VI 349
Db      33 GATRTVVYKVPREQPNTLIGSLADYGPVDVGHLYKLEVG---APYLRVDKGTGDI 87
Qy      350 RTRGPDREVEVSYQ-----LVTEASDQGRDGPRTTAVALFVSEVDNDNAPQ 398
Db      88 TTETSIDREGLRCONQLPDPCILEFEVSIITDLVONASPRLLBQOI--EVQDINDINTPN 145
Qy      399 FSEKRYVVOVREDVTPGAPVLRTVATSDRDKGSNAVVHYSIMSGNARGQFPLDQGTALDY 458
Db      146 FASPVITTLAIPENTNIGSLFPIPLASDRDAGPNVASYELQVAEDQ-----BEKQPOLIV 200
Qy      459 VSPIDVETTKETTLRVRAQDGRPPLSNVGLVTVQVLDINDNAPIFVSTPQATVLESV 518
Db      201 MGNIDREKRWSDYDLTIKVQDGGSPRA--TSALLRVTLVDINDNAPKERSYEAEISENS 259
Qy      519 PLGIVLVIAQALDADGNARLEFRLAGVGHDP-----FTINNGTGMISVAELDREE 572
Db      260 PIGHSVIOVKANDSOGANAEIETFP---HOAEVVRRLRLDNTGLITVQGPVDRED 315
Qy      573 VDFYSGVEARDHGTPALITASASVTVLVDVNDNPTF-----TOPEYTVRLNEDAA 624
Db      316 LSTLRPSVLAKDGTNPKSARAQVTVKMDNNAPIETIRGIGLVTHODGMANISEDA 375
Qy      625 VGTSVTVASAVRD--AHSVITYQITSGNTRNRPISITSGGG---LVSLALPLDYKL 677
Db      376 EETAVALLVQVSDRDEGENAAVTC--VVAGDVPOLRQASETSGDSKKKYFLQTTPLDYEK 434
Qy      678 ERQYVLAVTSADGR---QDTAQIVNVNTDANTHRPVQSHYTVANNEDRPACTTVLI 734
Db      435 VKDYTLIEIVAVDSGNPPLSSTNSLKVQVVDVNDNAPFTQSVTEVAPENNKKGEVIAEI 494
Qy      735 SATDEDTGENARITYFMEDSIPQ-----FRIDADTGAVTTOALEYDQVSYTLAITAR 788
Db      495 TADADSGSNALVYSLE---PEPAKGLFTISPETGEIOVKTSLDRQRESIELKVVA 551
Qy      789 DNGIPQKSDTTLVLELVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLOISATDRDGLNG 848
Db      552 DRGSPSLQGTATVLLVNLDCNDNDPKFMLSQNFVSNEMNPLSPVGMVTVIDGKGENA 611
Qy      849 RIVFYTGQGDGDDGPIVESTSGIVRTLRRLDRENAQVYLRAVADKGMPPARTMEVT 908
Db      612 QVQLSV---EQDNGDFVIOGTGTLISLSLSPRQOSTYTFOLKAVDGVGPPRSAYVGT 668
Qy      909 VTVLDVNDNPPVF--EODEFDVFEENSPIGLAVARTATPDDEGTAQIMYOIVEGNIP 966
Db      669 INVLDENDNAPYITAPNTSHKLLTPQTRIGETVSOVAEDPDSGNALIVISAGNPY 728
Qy      967 EVFQDLIFSGELTALVDLDYEDRPREVYLVIOQTSAPLVSRATVHRLDRNDNPPVLGN 1025

```

Db 729 GLFOIGSHSGAITEKIERRRHGLRLV-----VKVSDRG-KPRRYGTA 772

QY 1026 -FEILFNNVYNNRSSPPGGAIGRVPAH-----DPDISDLTY--SFERGNEI 1070

Db 773 LVHLVYNETLANRT-----LLETLLGHSIDTPIIDIDAGDEYERSKORGNIL 820

RESULT 6

US-08-268-161A-95

Sequence 95, Application US/08268161A

Patent No. 5798224

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/268,161A

FILING DATE: June 27, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Young J. Suh

REGISTRATION NUMBER: P-41,337

REFERENCE/DOCKET NUMBER: 27866/32149

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEO ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 1026 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-268-161A-95

Query Match 5.4%; Score 846; DB 1; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;

Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPPNANT-----LYRLGSGSGSPSEVEIDPRSG-VI 349

Db 33 GHATRVYKVPKPEQPPNTLIGSLAAYGPPDVGHLYKLEVG-----APILRVGKTKGDI 87

QY 350 RTGPGVDREVESEYQ-----LYEASPDQGRDPGPRSTTAAVFLSVEDNDNAQ 398

Db 88 TTETSIDREGIARCCQQLPGDPCILEFEVSIITDLYGNASRLLEGI--EVDQINNTN 145

QY 399 FSEKRYVOVREHVETGAPLRTASDRDKGSNAVVHYSIMSGNARGOTLLDAQCALV 458

Db 146 FASPIVTLTLPENTNGSLPEPIPLASDRGPGVASYELQVAEDQ-----EKKQPLTV 200

QY 459 VSPLDJETTKETTLRVARADGGRPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLSV 518

Db 201 MGNLDBERHDSYLTITKVDGSGPPRA--TSALIRVTLDTNDNAPFEPSPSYAEISENS 259

QY 519 PLGYLVLAHQALDADAGDNARLETRLAGVGHDP-----FTINNGTGISVAELIDREE 572

Db 260 PIGHSVIQVANDSDGANAIEIYTF-----HQAFVVRRLRLDRNTGILITQGVPRDSD 315

QY 573 VDFYSGVABRHGTALRASASVSTVLDVNDNNTF-----TQPEYIVRLNEDAA 624

Db 316 LSTLRFSVLAKORGTNPKSARAQVVTVTKOMNDNAPTIIEIRGILVTHQDMANISEDVA 375

QY 625 VGRSVYTVANVRD--ASHVITYQITSNTRRRFSITSGSGG-----LVSLALPLDYKL 677

Db 376 BETVALVGVSRDEGENAAVTC--VVAQVPPQLRQASSTGSKKKYFLQTTTPIIDYER 434

QY 678 ERQYVLAATVADQTR---ODTAQIVNVTANTHRPVPQSSHVTYVNVNEDRPAQTVVLI 724

Db 435 VQDYTIEIYAVDSGNPPLSTNSLKVQVVDVNDNAVFQSYTEVAFPERNNKGEVIAEL 494

QY 735 SATDEDTGNNARTTYEMEDSIPO-----FRIDADTGAVTQAELDYBDQVSTYLAITAR 788

Db 495 TASDADSGNNAELVSLF---PEPAKGFTISPETGEIQVTSILDRQRESEYELKVVA 551

QY 789 DNGIPEKSTTYLEILVNDVNDNAQPLDYSQSGSYEVEPPFTSYLQISATDRDGLNG 848

Db 552 DRGSPSLQGTATVVLVAVLDCNDNDPKFMLSQYFNSYMNMPALSPGVMTVTDGDKGENA 611

QY 849 RVFYTFQGGDDGDDPIVESTGIVRTTLARLDRENVAVYLAAYVDKGPAPARTMEYT 908

Db 612 QVQLSV---EQDNGDPVIONGTITLISLSPRQOSTTTPQLKAVDGGVPPRSAAVGT 668

QY 909 VTVLVDNDNPVF--EQDEFVFEENSPIGLAVARVATATDDEGTNAQIMTQIVEGNIP 966

Db 669 INVLDENDNAPYITAPNSHTKLLFPQTRLGETVSGVAEDSDGVNAELIYSIAGNIFY 728

QY 967 EYFQDIPSGELTALVDLDYERPEVLYTQATSAVLVRAVYHRLDRNDNPPLGK- 1025

Db 729 GLFOIGSHSGAITEKIERRRHGLRLV-----VKVSDRG-KPRRYGTA 772

QY 1026 -FEILFNNVYNNRSSPPGGAIGRVPAH-----DPDISDLTY--SFERGNEI 1070

Db 773 LVHLVYNETLANRT-----LLETLLGHSIDTPIIDIDAGDEYERSKORGNIL 820

RESULT 7

US-08-453-702A-95

Sequence 95, Application US/08453702A

Patent No. 5891706

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,702A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5891706and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 32657

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEO ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 1026 amino acids

TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-702a-95

Query Match 5.4%; Score 846; DB 2; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;  
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

```

QY 305 GYEVLTAVRATDGDAPNNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349
DB 33 GHATRVVTKVPEKQPPNTLIGSLADYGPDPVGHLYKLEVG-----APLRVDGKTGDI 87
QY 350 RTGPEVDEEVESYO-----LYVASDQGRDPGRSTTAAVFLSVEDNDNAPQ 398
DB 88 TETSIDREGLRQCQNLPGDPCILEPEVISTDLVQNASPRLEGGI--EVQDINDNFTN 145
QY 399 FSEKRYVQVREDVTPGAPVLRATASPRDKGSNAVYHYSIMSGNARQGFILDAQTALDY 458
DB 146 FASPVITTLAIPENTNIGSLFPPLASDRDAGPVGVAAYELQVABDQ-----EKKOPOLIV 200
QY 459 VSPLDYETTKETTLRVARADGGRPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESY 518
DB 201 MGNLDERMDSYDLTIKVDGSGPPRA--TSALLRVTYLDINDNAKFERPSYAELESENS 259
QY 519 PLGYLVIAVQALDADAGNARLERYLAGVGHDP-----FTINNGTGMISVAABLDREE 572
DB 260 PIGHSVIYQVANDSDQGANAEIETP-----HOAPEVVRRLRLDRNTGLITVQGPVDRD 315
QY 573 VDFYSGVEARHDGTPALTSASVSTYVDVNDNPTF-----TOPEYVRLMEDAA 624
DB 316 LSTLRPSVLAQRGTPKPSARQVAVTVKMDNDNAPTEIRIGIGLVTHODGMANISEVA 375
QY 625 VGTSVVTVASVDRD--AHSVITYQITSQNTNRNFSITSQSGG-----LVSLALPLDYKL 677
DB 376 EETAVALVQVSDRDEGENAAVTC--VVAADVPPQLQASBETSGSKKKYFLQTTPLDYEK 434
QY 678 EROQYLAVTASDGT--QDTAQIVNVTDANTHRPVFOSSHYYTVANNEDRPAGTTVLI 734
DB 435 VKQYTIETVAVDSGNPPLSSTNSLKVQVVDVNDNAPFTQSVTEVAFPENNKKEGVIAEI 494
QY 735 SATDEDTGENARITTFMEDSIPO-----FRIDADTGAVTTOALEYEDQVSTLAIAR 788
DB 495 TMSDADSGSNAELVSLD--PEPAKGLFTISPEVTEIOVKTSLPREQKSESTELKVVA 551
QY 789 DNGIPIKSDITTYLEILVNDVNDNAPOLFSDSYQGSYVEDVPFTSYLQISATDRDGLNG 848
DB 552 DRSPSLQGTATYLVAVLDCNDNDPKFMLSQVNFVSWENNIPALSPVGMVTVIDGDGENA 611
QY 849 RVPYTTQGGDGDGDFIVESTSGIVTTLRLDRENNVAQYVLAAYDKGMPARTMEYV 908
DB 612 QVQLSTV---EQDNGDFVITQNGTGITLSLSPDREQOSTYTFQLKAVDGVPFRSAAYGVY 668
QY 909 VTVLDVNDNPPV--BQDEPDEVVEENSPIGLAVARTADPDEGNNAQIMQIVGENTP 966
DB 669 INVLDENDNAPYTTASNTSHKLLTPQTRIGETVSQVAAEDPDSGVAAELIYSIAGNPNY 728
QY 967 EYFQDLIFSGELTALVDLDEDEPREVLYIQATSAPLVSRATVHVLRLDENDRPVLGN- 1025
DB 729 GLPQIGSHSGAIILEKIEIRRHGHLRLV-----VAVSDRG--KDPYRGTA 772
QY 1026 -FELFENNYTNRSSSPFGAIGRVPAH-----DPIISDLSLY--SFERGNEI 1070
DB 773 LVHLVYNETLANRT-----LLETLLGHSIDTPLDIDIDADPEYERSKORGNIL 820

```

# RESULT 8

US-09-639-95  
Sequence 95; Application US/0909639

Patent No. 6262237

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protoporpherin Materials and Methods

NUMBER OF SEQUENCES: 115

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago  
STATE: Illinois

COUNTRY: USA  
ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639

FILING DATE: 18 JUN 1998  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161

FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:

NAME: Greta E. No. 6262237 and  
REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1026 amino acids

TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-639-95

Query Match 5.4%; Score 846; DB 3; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;  
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

```

QY 305 GYEVLTAVRATDGDAPNNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349
DB 33 GHATRVVTKVPEKQPPNTLIGSLADYGPDPVGHLYKLEVG-----APLRVDGKTGDI 87
QY 350 RTGPEVDEEVESYO-----LYVASDQGRDPGRSTTAAVFLSVEDNDNAPQ 398
DB 88 TETSIDREGLRQCQNLPGDPCILEPEVISTDLVQNASPRLEGGI--EVQDINDNFTN 145
QY 399 FSEKRYVQVREDVTPGAPVLRATASPRDKGSNAVYHYSIMSGNARQGFILDAQTALDY 458
DB 146 FASPVITTLAIPENTNIGSLFPPLASDRDAGPVGVAAYELQVABDQ-----EKKOPOLIV 200
QY 459 VSPLDYETTKETTLRVARADGGRPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESY 518
DB 201 MGNLDERMDSYDLTIKVDGSGPPRA--TSALLRVTYLDINDNAKFERPSYAELESENS 259
QY 519 PLGYLVIAVQALDADAGNARLERYLAGVGHDP-----FTINNGTGMISVAABLDREE 572
DB 260 PIGHSVIYQVANDSDQGANAEIETP-----HOAPEVVRRLRLDRNTGLITVQGPVDRD 315
QY 573 VDFYSGVEARHDGTPALTSASVSTYVDVNDNPTF-----TOPEYVRLMEDAA 624
DB 316 LSTLRPSVLAQRGTPKPSARQVAVTVKMDNDNAPTEIRIGIGLVTHODGMANISEVA 375
QY 625 VGTSVVTVASVDRD--AHSVITYQITSQNTNRNFSITSQSGG-----LVSLALPLDYKL 677
DB 376 EETAVALVQVSDRDEGENAAVTC--VVAADVPPQLQASBETSGSKKKYFLQTTPLDYEK 434
QY 678 EROQYLAVTASDGT--QDTAQIVNVTDANTHRPVFOSSHYYTVANNEDRPAGTTVLI 734
DB 435 VKQYTIETVAVDSGNPPLSSTNSLKVQVVDVNDNAPFTQSVTEVAFPENNKKEGVIAEI 494
QY 735 SATDEDTGENARITTFMEDSIPO-----FRIDADTGAVTTOALEYEDQVSTLAIAR 788

```

Db 495 TASDADSGSNALVYSLE---PEPAKGLFTISPENGEIQVTSIDREORESEYELKVAA 551  
 Qy 789 DNGIPOKSDTLYLEILVNDVNDNAPOFLDSYQSGYVEDVPFTSVLQISATDRDGLNG 848  
 Db 552 DRGSPSLQGTATVAVLNDNDNDPKFMSLGFVNSFMENMPLSPGVAVTVIDGKXENA 611  
 Qy 849 RFEYFQGGDDDDGPIVSTSGIVRTLRRLDRNVAQVLAIVADKMPAPRTPEMYT 908  
 Db 612 QVQLSV---EQDNGDFVIONGTGTILSLSPDRQOSTYTFQLKAVDGGVPPRPAVYGV 668  
 Qy 909 VTLVDVNDNPPVFE---EODEDFVFEVENSPIGLAVARVATDPEDEGNAQIMQIVEGNIP 966  
 Db 669 INVLDENDNAPYITAPNSYSHLIPQTRLGETVSQVAABDDSGVNAELIYSIAGNIPY 728  
 Qy 967 EVFQLDIFSGELTALVDLYEDRPEVYVIOATSAPLVSRATVAVRLDRNDNPPVLGN- 1025  
 Db 729 GLFQIGSHGATILEKEIERRHGHLRLV-----VKVSDRG-KPFRYGYA 772  
 Qy 1026 -FEILFNMYVNNRSSSPFGAIGRPVPAH-----DPDISDSLTYY--SFERGNEL 1070  
 Db 773 LVHLVYNETLANRT-----LLETLLGHSIDTLPDIDDIADGPEYERSKQGNIL 820

## RESULT 9

PCT-US93-12588-95  
 ; Sequence 95, Application PC/TUS9312588  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 107  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ; ADDRESSER: Borun  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/12588  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/998,003  
 ; FILING DATE: 29 DEC 1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 31811  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 95:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1026 amino acids  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US93-12588-95

## Query Match

Best Local Similarity 29.1%, Pred. No. 1a-51;  
 Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

Qy 305 GYEVLTVAATGDAAPPNANI-----LYRLLESGSGSPSEVFEPRSG-VI 349  
 Db 33 GHATRVVVKVPEBOFPNTLIGSLADYGFPPDGHVYKLEFVG-----AAYLAVDAGTGDIF 87

Qy 350 RTRGVDRREEVSYQ-----LTVEASDGRDGPSTTAAVELSYEDDNDNAPQ 398  
 Db 88 TTERSIDREGAREQONQAPGDPCLLEFEVSTIDLVQNASPRLLSGQI--EVQIDINDTFN 145  
 Qy 399 FSEKRYVQVREDEYTPGAPVLRTASDRDKGSNAVHYSINSGNARQGFYLDAGTALDV 458  
 Db 146 FASPVITLAIPEENTNIGSLFPIPLASDRDAGPNQVASYELQVABDQ-----EEKQPOLIV 200  
 Qy 459 VSPLDYETTEKYLTVRAADGGGRPLNSVSGLVVQVYLDINDNAPIFSTPFGQATVESV 518  
 Db 201 MANDIREKWDSDYLTIKVQDGSPPRA-TSALLKVTYLDINDNAPKERPSYEAELSNIS 259  
 Qy 519 PLGYLVTHVQAIADADGNALLEYRLAGVGDPP-----FTINNGTGMISVALEIDREE 572  
 Db 260 PIGHSVIOVQKANDSDQGNAMIEYTF---HQAEVYRRLRLRNTGLIVQGVDRRED 315  
 Qy 573 VDFYSGVEARHDGTPALITASASVSVTVLVNDNPPF-----TQPEYTVRLMEDAA 624  
 Db 316 LSTRFSTLADRGNTNPKSARAQVAVVVKOMNDNAPTEIRIGILVTHQDMANISEDA 375  
 Qy 625 VGTSVTVSAVDRD--AHSVITYQITGNTRNRSITSGSGG-----LVSLALPLVYKL 677  
 Db 376 ERTAVVALVQSDRDEGENAAVTC-VAGADVPPOLRQASETSQSKKYPLOTTPPLYEK 434  
 Qy 678 EROYVLAVTASDGR---ODTAQIVAVNTDANTERPVSQSHYTVANNEDRPAGTAVLI 734  
 Db 435 VKDTIRIVANDSGNPLSTNSLSKQVAVVNDNAPFTQSVTEVAPPENKPEVIAEI 494  
 Qy 735 SATDEDTGENARITYPMEDSIPO-----FRIDAGVAVTQAEVDYEDQVSYLTAITAR 788  
 Db 495 TASDADSGSNALVYSLE---PEPAKGLFTISPENGEIQVTSIDREORESEYELKVAA 551  
 Qy 789 DNGIPOKSDTLYLEILVNDVNDNAPOFLDSYQSGYVEDVPFTSVLQISATDRDGLNG 848  
 Db 552 DRGSPSLQGTATVAVLNDNDNDPKFMSLGFVNSFMENMPLSPGVAVTVIDGKXENA 611  
 Qy 849 RFEYFQGGDDDDGPIVSTSGIVRTLRRLDRNVAQVLAIVADKMPAPRTPEMYT 908  
 Db 612 QVQLSV---EQDNGDFVIONGTGTILSLSPDRQOSTYTFQLKAVDGGVPPRPAVYGV 668  
 Qy 909 VTVVDVNDNPPVFE---EODEDFVFEVENSPIGLAVARVATDPEDEGNAQIMQIVEGNIP 966  
 Db 669 INVLDENDNAPYITAPNSYSHLIPQTRLGETVSQVAABDDSGVNAELIYSIAGNIPY 728  
 Qy 967 EVFQLDIFSGELTALVDLYEDRPEVYVIOATSAPLVSRATVAVRLDRNDNPPVLGN- 1025  
 Db 729 GLFQIGSHGATILEKEIERRHGHLRLV-----VKVSDRG-KPFRYGYA 772  
 Qy 1026 -FEILFNMYVNNRSSSPFGAIGRPVPAH-----DPDISDSLTYY--SFERGNEL 1070  
 Db 773 LVHLVYNETLANRT-----LLETLLGHSIDTLPDIDDIADGPEYERSKQGNIL 820

## RESULT 10

PCT-US95-08071-95  
 ; Sequence 95, Application PC/TUS9508071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ; ADDRESSER: Borun  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; PCT-US95-08071-95







```

; Sequence 103, Application US/08453695A
; Patent No. 5708143
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-103

Query Match          5.4%; Score 846; DB 1; Length 1203;
Best Local Similarity 29.1%; Pred. No. 1.3e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPNNANI-----LYRLLESGSGSSEVFETIDPSG-VI 349
DB 33 GHATRVVYVPEBQPNITLIGSLADYGPDDVGHLYKLVG-----APLRYDGKTDIF 87
QY 350 RTGSPVDRREVSQ-----LYEASDQGRDPGRSTTAAPLVSBDNDNAPQ 398
DB 88 TETSTIDREGIARECQQLPEDPCILEPEVSTIDLVGNAPRLLEGI--EVQDINDNTN 145
QY 399 FSEKRVVQVREDVTPGAPVLRVTASDRDKGSNAVVHYSIMSGNAGQFYLDAGTALDV 458
DB 146 PASPVITLAIPEWNTIGSLFPIPLASDRAGPVGASVELQVAEDQ-----SEKQQLIV 200
QY 459 VSPDLVETTKETYLKRAQDGGRRPLSNVGLVTVQVLDINDNAPLFVSTPPQATVLESV 518
DB 201 KGNLDREERDSYDLTKVQDGSPPRA--TSALLRVTLTDNDNAPKFEPSYEAEISENS 259
QY 519 PLGYLVIAVOAIDADAGDNARLEYRLAGVCHDFP-----FTINNGTGMISVAABEDRE 572
DB 260 PIGHSVIQKANDSDGANAIEYTF---HQAPEVVRRLRLDRNTGLITVQGPVDRD 315
QY 573 VDEYSFGVEARDHGTALTNASASVTVLDVNDNPPF-----TOPEYTVRLNEDAA 624
DB 316 LSTLRFSVLAKRGTPKSAFAQVVTVVQMDNNAFTIEIRIGLIVTHQDGMANISEDA 375
QY 625 VGSVTVTASAVDD--AHSVITQITSGNTRNRFSTSGSGS-----LVSLALPDIYCL 677
DB 376 EETAVLALVOVSDDBEGNAAVTC--VVAAGVPPQLRASSETGSDSKKTYFLQTTPIDYK 434
QY 678 EROVYLAVTASDSTR---OPTAQIVNVNTDANTHRFVFOSSHYYTVVNNEDRPAGTVVLI 734

```

```

DB 435 VKDYTIEIVADSGNPPLSTNSLKQVVDVNDNAPVFTQSTVEAFENKPGEVIAEI 494
QY 735 SATDEDTGENARITYFMEDSIPO-----FRIDADTGAVTTOAEILDYDOVSYTLTAIR 788
DB 495 TNSDADSSNAELVYSLE---PEPAKGLFTISSETGHIQKTSIDRQRSSEYELKAVAA 551
QY 789 DNGIPQKSDITYLEILVNDVNDNAPQFLRDSYQGSVEVDVPPFTSVLOISATDRDSGLNG 848
DB 552 DRGSPSLQGTATVAVNVLDNDNDNPKFMLSQYNFSVNMENMALSFGVAVTVIYDGDGENA 611
QY 849 RVFTYFGGDDGDDDFVYESTSGIVRTLRDLRENNVAYVLRARAIVDKGMPAPRPMVTV 908
DB 612 QVQSV--EQDNDGPVYQNGTGTLLSSLPDRQQSTYTFQLARAVDGVPPRSAYVGT 668
QY 909 VTVDVNDNPPVF--EODEPVPVEENSPIGLAVARVATPDDEGTNAQINQIVEGNIP 966
DB 669 INVLDENDNAPYITAPSTSHKLITPQRLGETYSQVAAEPPDGSVNAELYSTAGNIPY 728
QY 967 EVFQIDIFSGELVAVLDIYEDRPEYVLVIQATAPLVSRAVTVHRLIDRNDNPPVIGN- 1025
DB 729 GLFQIGSHSGAITTEKEIERHHGLHRLV-----VKVSDRG-KPPRYGTA 772
QY 1026 -FEILFNMYVNNRSGSPFGCAIGRVPAH-----DPOISDSLTY--SPERGNEL 1070
DB 773 LVHLVNETLANRT-----LLETLLGHSIDTPIDIDLAGDPEYERSKQGNIL 820

```

```

RESULT 14
US-08-268-161A-103
; Sequence 103, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-268-161A-103

Query Match          5.4%; Score 846; DB 1; Length 1203;
Best Local Similarity 29.1%; Pred. No. 1.3e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPNNANI-----LYRLLESGSGSSEVFETIDPSG-VI 349

```

```

Db      33 GHATRVVVKVBEOPNTLIGSLADYGFPPDVGHLYKLEVG-----APYLKVDKGTGDI 87
Qy      350 RTGPGVDRREEVESYQ-----LTVEASDQGRDPCGPRSTTAAVFLSVEDNDNAPQ 398
Db      88 TTETSIDREGARECONQJPCILIEFEVSITLTVGNASFLLEBQI--EVQDINDNTN 145
Qy      399 FSEKRYVQVREDVTGAPVLRTASDRDKGSNAVHYISMSGNARQGFYLDAGTALDV 458
Db      146 FASPVITLAIPEENTNIGSLFPIPLASDRDAGPNNGVASYELQVAEDQ-----BEKQPLI 200
Qy      459 VSPIDYETTKETYLRLVRAODGRPPLSVNGLVTVQVLDINDNAPIFVSTPQATVLESV 518
Db      201 MGNLDRERWDSYDLTKVQDGSPPRATSAALRLVTLDTNDNAPKFERPSEYAELESENS 259
Qy      519 PLGYLVLTVOAIDADAGDNARLEERLAGVGHDP-----FTINNGTGMISVAALDREE 572
Db      260 PIGHSVIOVKANDSDGANAIEITYF---HQAPEVVRRLRLDRNTGLITVQGPVDRD 315
Qy      573 VDFYSFGVEARDHGTALRTASASVTVLVDVNDNPTF-----TQPEYTVRLMEDAA 624
Db      316 LSTRFSTVLAKDRGTNPKSARAQVVTVKDMNDNAPTEIRIGLGLVTHQGMANISEDVA 375
Qy      625 VGTSVTVASANDRD--AHSVITYQTSGTNRKRSITSGSGG-----LVSLALPLDYK 677
Db      376 EETVALVQVSDRDEGENAAVTC--VVAAGDVPPQLRQASSETSDSKKXYLTQTPPLDYEK 434
Qy      678 EROQYLAVTASDGR---QDTAQIVNVNTDANTHRPVSQSHYTVNVEDRPACTTVLI 734
Db      435 VKQDTIEIVAVDSGNPPLSTNSLKVQVVDNDNAPFTQSVTEVAPENNKGEVIAEI 494
Qy      735 SADEDTGENARITYFMECSIPO-----FRIDATGAVTTOAEILDYEDQVSYTLAITAR 788
Db      495 TASDADSGSNAELVYSLE---PEPAKGLFTISPETGEIQVTKSLDREQRSYELKVAA 551
Qy      789 DNGIPOKSDTYLEILVNDVNDNAPQPLADSYQSYVEDVPPFTSVLQISATDSDGLNG 848
Db      552 DRGSPSLQGTATVAVNLDCNDNDPKFMLSINFSVNMNPALSPVGMVTVITDDKEMA 611
Qy      849 RVFTYFGQDGDGDFIVESTSGIVRTLRLDRRENVQVYLRAVADKMPPARTMEVT 908
Db      612 QVQLSV---EQDNDGVYQNGTGTILSLSPDEQSGSTYTFQLAKADGCVPPRSAYGVY 668
Qy      909 VTVLDVNDNPPVF--EQDEPDVFEENSPIGLAVARTATDPDEGTNAQIMYQIVBGNIP 966
Db      669 INVLDENDNAPYITAPSNTHSKLLTPQRLGETVSGVAAEDPDGVAELIYSIAGNPFY 728
Qy      967 EVFPLDJFSGELTALVLDYDRPEYVLVYQATSAPLVSRATVAVRLDRDNDPPVLGN- 1025
Db      729 GLFQIGSHSGAITLEKEIERHHGILHRLV-----VKVSDRG--KEPRYGT 772
Qy      1026 -FELIPNNVYVTRSSSPFGAIGRVAH-----DPDISDLTY--SPERGNEL 1070
Db      773 LVHLYVNETLANR-----LLETILGHSUDTPDIDIDAGPPEYERKQKGNIL 820

```

RESULT 15  
 US-08-453-702A-103  
 : Sequence 103, Application US/08453702A  
 : Patent No. 5891706

: GENERAL INFORMATION:  
 : APPLICANT: Suzuki, Shintaro  
 : TITLE OF INVENTION: Protocadherin Materials and Methods  
 : NUMBER OF SEQUENCES: 115  
 : CORRESPONDENCE ADDRESSES:  
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 : STREET: 233 South Wacker, 6300 Sears Tower  
 : CITY: Chicago  
 : STATE: Illinois  
 : COUNTRY: USA  
 : ZIP: 60606

COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/453,702A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5891706and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 32657
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELETYPE: 25-3856
: INFORMATION FOR SEQ ID NO: 103:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1203 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-453-702A-103

Query Match      5.4%; Score 846; DB 2; Length 1203;
Best Local Similarity 29.1%; Pred. No. 1,3e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

Qy      305 GYEVLYTRARDGAPPRANI-----LTRLGSGSGSPBEVEIPIRSG-VI 349
Db      33 GHATRVVVKVBEOPNTLIGSLADYGFPPDVGHLYKLEVG-----APYLKVDKGTGDI 87
Qy      350 RTGPGVDRREEVESYQ-----LTVEASDQGRDPCGPRSTTAAVFLSVEDNDNAPQ 398
Db      88 TTETSIDREGARECONQJPCILIEFEVSITLTVGNASFLLEBQI--EVQDINDNTN 145
Qy      399 FSEKRYVQVREDVTGAPVLRTASDRDKGSNAVHYISMSGNARQGFYLDAGTALDV 458
Db      146 FASPVITLAIPEENTNIGSLFPIPLASDRDAGPNNGVASYELQVAEDQ-----BEKQPLI 200
Qy      459 VSPIDYETTKETYLRLVRAODGRPPLSVNGLVTVQVLDINDNAPIFVSTPQATVLESV 518
Db      201 MGNLDRERWDSYDLTKVQDGSPPRATSAALRLVTLDTNDNAPKFERPSEYAELESENS 259
Qy      519 PLGYLVLTVOAIDADAGDNARLEERLAGVGHDP-----FTINNGTGMISVAALDREE 572
Db      260 PIGHSVIOVKANDSDGANAIEITYF---HQAPEVVRRLRLDRNTGLITVQGPVDRD 315
Qy      573 VDFYSFGVEARDHGTALRTASASVTVLVDVNDNPTF-----TQPEYTVRLMEDAA 624
Db      316 LSTRFSTVLAKDRGTNPKSARAQVVTVKDMNDNAPTEIRIGLGLVTHQGMANISEDVA 375
Qy      625 VGTSVTVASANDRD--AHSVITYQTSGTNRKRSITSGSGG-----LVSLALPLDYK 677
Db      376 EETVALVQVSDRDEGENAAVTC--VVAAGDVPPQLRQASSETSDSKKXYLTQTPPLDYEK 434
Qy      678 EROQYLAVTASDGR---QDTAQIVNVNTDANTHRPVSQSHYTVNVEDRPACTTVLI 734
Db      435 VKQDTIEIVAVDSGNPPLSTNSLKVQVVDNDNAPFTQSVTEVAPENNKGEVIAEI 494
Qy      735 SADEDTGENARITYFMECSIPO-----FRIDATGAVTTOAEILDYEDQVSYTLAITAR 788
Db      495 TASDADSGSNAELVYSLE---PEPAKGLFTISPETGEIQVTKSLDREQRSYELKVAA 551
Qy      789 DNGIPOKSDTYLEILVNDVNDNAPQPLADSYQSYVEDVPPFTSVLQISATDSDGLNG 848
Db      552 DRGSPSLQGTATVAVNLDCNDNDPKFMLSINFSVNMNPALSPVGMVTVITDDKEMA 611
Qy      849 RVFTYFGQDGDGDFIVESTSGIVRTLRLDRRENVQVYLRAVADKMPPARTMEVT 908
Db      612 QVQLSV---EQDNDGVYQNGTGTILSLSPDEQSGSTYTFQLAKADGCVPPRSAYGVY 668
Qy      909 VTVLDVNDNPPVF--EQDEPDVFEENSPIGLAVARTATDPDEGTNAQIMYQIVBGNIP 966

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 17:44:36 ; Search time 22139 Seconds  
(without alignments)  
5401.276 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 15545  
Sequence: 1 MMSPTGVLPPTPPPLLL.....AGTVDEDSGSEFLFFNLH 2923

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DBV=x1h  
-O=/cgn2\_1/USPTO.spool/US09916849/runat\_11022004\_143816\_20938/asp\_query.fasta\_1.3079  
-DB=GenEmbl -QFMT=fastsp -SUPRFX=rge -MINMATCH=0 -LOOPEXT=0 -COOPEXT=0  
-INITs=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09916849@cgn2\_1 15745 @runat\_11022004\_143816\_20938 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_nu:\*  
21: em\_ov:\*  
22: em\_or:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	15545	100.0	8772	6 AX224128	AX224128 Sequence
2	15545	100.0	10531	6 AX549238	AX549238 Sequence
3	15545	100.0	10531	6 AP234887	AP234887 Homo sapi
4	15518.5	99.8	8871	6 AX224126	AX224126 Sequence
5	15279	98.3	9321	6 AX367102	AX367102 Sequence
6	14647.5	94.2	8950	10 AB028499	AB028499 Mus muscu
7	12833	82.6	8924	9 DB7469	DB7469 Human mRNA
8	12339	79.4	24370	6 AX646739	AX646739 Sequence
9	12339	79.4	169241	9 AB065955	AB065955 Homo sapi
10	12339	79.4	169241	2 AL390252	AL390252 Human DNA
11	12339	79.4	227194	2 BX284647	BX284647 Homo sapi
12	12001	77.2	235320	2 AC106175	AC106175 Rattus no
13	12001	77.1	235327	2 AC095463	AC095463 Rattus no
14	11883	77.1	254975	2 AC095125	AC095125 Rattus no
15	11892	76.5	248059	2 AC113756	AC113756 Rattus no
16	11783.5	75.8	89129	10 AL672200	AL672200 Mouse DNA
17	11097	71.4	6699	10 AB011529	AB011529 Rattus no
18	10517.5	67.7	234932	2 AC093365	AC093365 Mus muscu
19	9016	58.0	10919	10 AF031572	AF031572 Mus muscu
20	8991.5	57.8	11660	2 AL157901	AL157901 Homo sapi
21	8781.5	56.5	9045	6 AX223857	AX223857 Sequence
22	8781.5	56.5	11389	6 AX549158	AX549158 Sequence
23	8781.5	50.4	11389	6 AP231024	AP231024 Homo sapi
24	8736.5	56.2	9087	6 AX226507	AX226507 Sequence
25	8710	56.0	10195	6 AE0094	AE0094 Sequence 4
26	7836.5	50.4	11956	9 AF231023	AF231023 Homo sapi
27	7836.5	50.4	11965	6 AX549370	AX549370 Sequence
28	7819.5	50.3	12348	6 AX662321	AX662321 Sequence
29	7780.5	50.1	11868	10 AB011528	AB011528 Rattus no
30	7766	50.0	11648	6 AX174748	AX174748 Sequence
31	7736.5	49.8	11541	10 AF427498	AF427498 Mus muscu
32	5554	35.7	206908	9 AC121252	AC121252 Homo sapi
33	5457	35.1	88326	9 AC005923	AC005923 Homo sapi
34	4984.5	32.1	239657	2 AC096455	AC096455 Rattus no
35	4825	31.0	12634	3 AF172329	AF172329 Drosophi1
36	4783.5	30.8	12200	3 AB028498	AB028498 Drosophi1
37	4766.5	30.7	143393	2 AC015222	AC015222 Drosophi1
38	4766.5	30.7	182494	3 AC009343	AC009343 Drosophi1
39	4766.5	29.9	182494	3 AE003828	AE003828 Drosophi1
40	4651.5	29.7	257909	9 AE0091437	AE0091437 Homo sapi
41	4584.5	29.5	6791	6 AE0092	AE0092 Sequence 2
42	4453	28.6	2653	10 AF031573	AF031573 Mus muscu
43	4433	28.5	3343	10 BC005499	BC005499 Mus muscu
44	3769.5	24.2	3912	6 AC098223	AC098223 Sequence
45	3602.5	23.2	216773	2 AC116764	AC116764 Mus muscu

ALIGNMENTS

RESULT 1

AX224128  
LOCUS AX224128 8772 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 3 from Patent WO0161003.  
ACCESSION AX224128  
VERSION AX224128.1 GI:15554400  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 Testa, T.T.  
The human homologue of flamingo  
Patent: WO 0161003-A 3 23-AUG-2001;  
JOURNAL SWITZERLAND BEECHAM PLC (GB)  
FEATURES  
source  
1. 8772  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 1696 a 2814 c 2564 g 1698 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 8772  
Score: 15545.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-916-849a-3 (1-2923) x AX224128 (1-8772)  
QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
Db 1 ATGCGAGGCCGCGCCACCGCGCTCCCTCCCAAGCCGCGCGCGCTGCTGCTG 60  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
Db 61 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyThrLeuCysProSer 60  
Db 121 GGGTCCAGGGGAGAGAGCTCTTGGGGGGCTGGCGCCCAATGAGCTGGCTCTGCATCC 180  
QY 61 SerAlaSerAsnLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
Db 181 TCAAGCTCGAAGCTCTGCTGCTTACACAGCGCTGACAGGATCGGGCACTGAGTGA 240  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGlnAlaHis 100  
Db 241 GGGCACTGGTACCCCAACAGATGAGCTGAGGGTGGTGGTCCAGATCCGAGGCCAT 300  
QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
Db 301 ATTCCCTTACACAGCTCTGCTGAGGGCTGCTGGAGCTGCTGCTGCTGCTGCTGCT 360  
QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGluGlnHisProCysLeuValAla 140  
Db 361 GGGCACTTTCCTCCCAAGGGGAGCTCACTGCCAGAGAGACCCCGTCTTAAAGGCT 420  
QY 141 ProArgLeuArgCysGlnSerCysValLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 421 CCAAGGCTCAGATGCCAGCTCTGCTGAGAGCTGCAAGGCCCCCGGCTCAAGGGAGAA 480  
QY 161 ArgSerProGluGlnSerLeuGlyLysArgValArgValAsnThrAlaProGln 180  
Db 481 AGGTACCAAGAGAGTCTCTGGTGGGCTGCGAAGAGAAATGTAATACACCCCCAG 540  
QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
Db 541 TTCACAGCCCCCAGCTACAGGGCACAAGTCCCGAGAGAACAGGACAGGCAACCCCTGTT 600

QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGlnAlaGlyArgLeuGluTyrThrMet 220  
Db 601 GCATCCCTGAGGGCCATGATGCCGAGCAGGGTGAAGGAGTGCATGAGATGACACAG 660  
QY 221 AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla 240  
Db 661 GATGCCCTCTTGTATGAGCCGCTCCCAACAGTCTTCCCTGAGACCACTGATGCA 720  
QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
Db 721 GTTACCAAGCCAGAGAGCTGATGTAGACCAAGAGCACCGTCTTCAAGGGTCAAG 780  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
Db 781 GGGCAGAGACCAAGGATGCCGAGAGTGCCTGGCTGACACTCAACATCTTGGTACT 840  
QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGlu 300  
Db 841 GACACCAATGACCAATGACCTGTGTTCAGCAGCAGAGTACAGAGAGGCTCAGGGAG 900  
QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 901 AACCTGAGAGTTGGCTATGAGGTGCTCACTGTCAGGGCCACGATGTGATGCCCTCC 960  
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGluValPhe 340  
Db 961 AATGCCAATATTCTGTACCGCTGCTGAGAGGGGTCTGGGGGGGAGCCCTGTAAGCTTT 1020  
QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
Db 1021 GAGATCGACCTTCGCTGCTGGGTGATCCGACCCCTGTCCTGATCGGAGAGGGTG 1080  
QY 361 GluSerTyrGlnLeuThrValGluValAspAspGlnGlyArgAspProGlyProArgSer 380  
Db 1081 GATCTTACACCTACCGGTAGAGGACAGTACAGGCTCGGAGCCCGGCTCTCGAGT 1140  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAlaProGlnPheSer 400  
Db 1141 ACCACAGCCGCTGTTTCTCTTCTGTGAGAGATGACATGATTAATGCCCCCAAGTTAGT 1200  
QY 401 GluLysArgTyrValValGluValArgGluAspValThrProGlyValProValLeuArg 420  
Db 1201 GAGAAAGCTGATGTGTCCAGGTGAGGAGATGATGACTCCAGGGGCCCTCACTCCGA 1260  
QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSerIleMetSer 440  
Db 1261 GTCACAGCTCGAGTCGAGAGACAGGGAGCAATGCCGTGTGCTCATTAATGATCTGAGT 1320  
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValLeuAspValValSer 460  
Db 1321 GGCATATGCTCGGGGAGCAGTTTATCTGATGCCAGACTGGAGCTCTGATGTGTGAGC 1380  
QY 461 ProLeuAspTyrGluThrThrLysGluTyrThrLeuArgValAlaArgAlaAspGly 480  
Db 1381 CCTCTTGAATATGAGACAGACAGAGAGTACCCCTACGGGTGCGAGACAGAGATGTGGC 1440  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1441 CGTCCCACTCTCAATGTCTCTGCTGCTGAGACATGATCTCTGATATCAACAGC 1500  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
Db 1501 AATGCCCCCATCTTGTGTGAGACACCCCTTCCAGGGCTACGTCTCTGAGAGGCTCCCTTA 1560  
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1561 GGTACCTGTGTTCTCAATGTCTCAGGCTATCAAGCTGATGCTGTGTGACATGCCCGCTG 1620  
QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1621 GAATACCGCTTGTGCTGGGTGGAGCATGACTTCCCTTCAACATGACATGACAGGC 1680  
QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580



```
Db 1681 TGGATCTCTGGCTGCTGAACTGGACCGGGAGAGTTGATTTCTACAGCTTTGGGGTA 1740
Qy 581 GIuAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1741 GAAGCTCGAGAGCAATGGACCTCGACACTGCTGCTGGCCAGAGTCTGAGCTGACTCTC 1800
Qy 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db 1801 CTGGATGTCACAGAACATCCAACTTTACCCCAACAGATACAGTGGCGCTCAAT 1860
Qy 621 GIuAspAlaAlaValGlyThrSerValValThrValSerAlaValAspAspAspAlaHis 640
Db 1861 GAGATGCACTGTGGGCAACGAGCTGGTGAAGGTGACGCTGAGCTGAGACCTGATGCTCAT 1920
Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db 1921 AGTGCATCACCTTACAGATCACAGTGGCAATATCTCGAAACCGCTTCTCCATCCAGCC 1980
Qy 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnArgGln 680
Db 1981 CAAGTGTGGTGGCTGGTATCCCTTGCCTGCCACTGSACTGACAACTTGAGCGGCGAG 2040
Qy 681 TyrValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db 2041 TATGTGTGGCTGTATACGCGCTCCGATGGCACTCGCGAGACACGCGACAGATGTGGTG 2100
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 2101 AATGTCACGAGCGCAACCAACCATGCTCTGCTTTCAGAGCTCCCACTATACAGTGAT 2160
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAsp 740
Db 2161 GTTATATAGGAGCCGCGCGGAGGACCAACGAGTGGCTGATCAGCGCCACGATAGAGAC 2220
Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACAGGTGAGATGGCCGCACTACCTTCATGAGAGAACAGCATCCCAAGTTCGCGATC 2280
Qy 761 AspAlaAspThrArgAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db 2281 GATGAGACACGGGGGCTGTCAACCAACGCTGAGCTGAGCTAGAGACCAAGTGTCT 2340
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db 2341 TACACCTCGGCACTTACGCTCGGACATGGCATTCCTCCAGAGTCCGACCACTCAC 2400
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db 2401 CTGGAGATCTCGGTGACGACGTGATATACAAATGCCCTCAGTTCTCGAGACTCTCAC 2460
Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGCGCATGTCTATGAGAGATGTGCACCTTCACTAAGCTGTGACATCTAGCCACT 2520
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860
Db 2521 GATCGTGAATCTGGACTTATATGGAGGGGCTTCTTCACTTCCAAAGGAGGCGAGATGA 2580
Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GACGGTGACTTATATGTGAGTCCACGTGAGCACTGCGCAACCTTACGAGAGCTGAT 2640
Qy 881 ArgGluAsnValAlaGlnTyrThrValLeuArgAlaTyrAlaValAspGlyGlyMetProPro 900
Db 2641 CGAGAGAACGTGGCCAGATATGTCTTGGGGCATATGACAGTGGACAAAGGAGTGCCCA 2700
Qy 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db 2701 GCCCGCACCACTTATGAGAGTGAAGTCACTGTGTGATGAGAAATGCAATCCCTCTCTC 2760
Qy 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTTGACAGAGATGAGCTTATGATGTCTTTGTGSAAGAGAACACCCCAATTGGGCTAGCCGTG 2820
Qy 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db 2821 GCCCGGTCAAGCACTAGCACTGACCCCGAGTGAAGGACCAATGCCAGATTATACAGATT 2880
Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla 980
Db 2881 GTGAGGGCAACATCCCTGAGGTCTTCCAGCTGACATCTTCCGGGGAGACTACAGCC 2940
Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGGTGACTTATGACTACAGAGACCGGCTGATGATCGCTGTGATCACAGGCCACGTCA 3000
Qy 1001 AlaProLeuValSerAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020
Db 3001 GCTCCTCTGGTGAACCGGGGCTACAGTCCAGTCCGCTCTTTCAGCCGAATGACAACTCA 3060
Qy 1021 ProValIleuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040
Db 3061 CCAGTGTGGGCACTTGAATCTTTTCAACAACATATGTCAACAAATCGCTCAAGCAGC 3120
Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3121 TTCCCTGGGGGTGCATTTGGCCGATACCTGCCATACCTGTATATCTCAGATAGTCTG 3180
Qy 1061 ThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db 3181 ACTTACAGCTTTGAGCGGGGAATGAATCAAGCTGTCTCTGCTCAATGCTCCACGGGT 3240
Qy 1081 GlnLeuIleLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal 1100
Db 3241 GAGCTGAAGCTAACCCCGCACTGAGCAACACCGGCTCTGAGGCGCATCATAGCGTG 3300
Qy 1101 LeuValSerAspGlyValHisSerValThrAlaGlnGlyValLeuArgValThrIleIle 1120
Db 3301 CTGGTGTACAGCGGCTGACACAGGTACCGCCACAGTGGCGGCTGAGTACCATCATTC 3360
Qy 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140
Db 3361 ACCGATAGATAGTCAACCAACAGATCAAGCTGTGGAGAGCAGTCAACCGAGCGC 3420
Qy 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160
Db 3421 TTCTGTACACCACTGACGCTTCTTATCAGCGGTGGCGGCGGCAAGCCGCA 3480
Qy 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180
Db 3481 CCGGACCACTGGTGTGCTTCAACGTACAGCGGACCAACGCCCCCGGGGGCCACATC 3540
Qy 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu 1200
Db 3541 CTCAACGTGAGCGTGTGGTGGGCGACCGCGCGGGGGCGGGGGCGGCTTCTCTG 3600
Qy 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220
Db 3601 CCCCTTAGGACCTCGCAGGAGCGGCTTATACCTCAACCGCAGCTGTGACGCGCATCTCG 3660
Qy 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240
Db 3661 GCACAGCGGTGCTGCGCTTCAAGACACACATCTGCTGGAGAGCCCTCGAAGACATAC 3720
Qy 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260
Db 3721 ATGGCTGCGGTGCGGTGCTGCGCTTGAATCCCTCGCGCTTCACTGCGCTCTCTCC 3780
Qy 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280
Db 3781 GTGCTCTTCCGGCCCACTCACCCGCTCGAGGGCTGCGCGGCTCCGCGGCTTC 3840
Qy 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300
Db 3841 ACGGGTACTATCTGCGAGACCGAGGTGACCTTCTGATCTCGCGGCTCTGTGGCCCCAC 3900
```

QY 1301 G1YArgCysArgSerArgGluGlyGlyTYrThrCysLeuCysArgAspGlyTYrThrGly 1320  
 DB GGCGCGTGGCGGAGCGCGGAGGGGGGCTACACCTGCTCTGTGTATGGCTACAGGGGT 3960  
 QY 1321 G1uHicGysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIysAsnGly 1340  
 DB 3961 GAGCAGCTGTAGGTGAGTGTCTGCTCCAGCGCTTGCGACCCCGGGGTGTCTGCAAGAAATGGG 4020  
 QY 1341 G1YThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysProSerGlyAspPhe 1360  
 DB 4021 GGACCTGTGTCAACTGCTGTGTGGGGGCTTCAAGTGCATGTGGCACTTGGAGACTTC 4080  
 QY 1361 G1uIysProTYrCysGluValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 DB 4081 GAGAGAGCCCTACTGCGAGGTGACCAAGCGAGCTTCCCGCCCACTCTTCACTCACTTT 4140  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIysGlyArg 1400  
 DB 4141 CGGGCCCTGGCGAGCGTTTCACTTCACTGAGCCCTCTCTGTGGCACAAGAGAGCGC 4200  
 QY 1401 AspGlyLeuLeuLeuTYrAsnGlyArgPheAsnGlyIuIysHisAspPheAlaLeuGlu 1420  
 DB 4201 GAGGGGTGTCTGTGTGACATGGGCGCTTCAATGAGAGCATGACTTGTGGCCCTCGAG 4260  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyIuSerThrThrValSer 1440  
 DB 4261 GTGATCAGAGAGAGGTCCAGCTCACTTCTGCGAGGGAGTCAACACCAAGGTGTCC 4320  
 QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuIysTYr 1460  
 DB 4321 CCATTGTGTGCGAGAGAGTCACTGATGAGGCACTGAGTACGATGCGCTGAATCTTAC 4380  
 QY 1461 AsnIysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIysValAla 1480  
 DB 4381 AAATACCACTGTGTGGTCAAGAGGCTCCCAAGGCGCCCAAGAGCAAGAGAGTGGCT 4440  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB 4441 GTGTGATCCGTGATGGCTGTGACACAGAGTGGCTTGGCTTGGATGTCTGGGG 4500  
 QY 1501 AsnTYrSerCysAlaIleAsnGlyThrGlnGlyIysSerIysIysSerLeuAspLeuThr 1520  
 DB 4501 AACTACTCCGTGTCTGCGAGGCGCCCAAGGAGTGGCGAAGAAAGTCTGAGATCTGAGC 4560  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGlyIysPheProValArgMetArg 1540  
 DB 4561 GGGCCCTGTGTAAGGGCGGGGTGCTGACCTGCCAGAGCTTCCAGTCCGAATGCGG 4620  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 DB 4621 CAGTGTGTGGGTGACATGCGGAACCTGCAAGTGAACGCGGCACTTACAGCTGAC 4680  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIysIysAsnValCysAspSer 1580  
 DB 4681 TTATATGCCAACAATGACCGTGTGCTGCGCTGCCCAAGAAAGACGTGTGTGACAGC 4740  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
 DB 4741 AACACTTGGCCACATGGGGGCACTGTGTGAACAGTGGGACCGCTTCACTGAGCTCG 4800  
 QY 1601 ProLeuGlyPheGlyGlyIysSerCysAlaGlnIleMetAlaAsnProGlnHisPheLeu 1620  
 DB 4801 CCCCTGGGCTTTGGGGGCAAGAGCTGCGCCCAAGAAATGCCCAATCAGCAGCACTTCTG 4860  
 QY 1621 GlySerSerLeuValAlaThrPheIleGlyLeuSerLeuProIleSerGlnProTYrIleu 1640  
 DB 4861 GGAGAGAGCTGTGGGCTGTGGCATGGCTTCTCGCTGCCATCTTCCCACTGTATCTC 4920  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 DB 4921 AGCCTCATGTTCGCGAGCGCCAGGCGGACGAGTGTCTGTGAGGCGCATCACAGGGGG 4980

QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 DB 4981 CGCAGACCATATCACTCTACGTTACAGAGGGCCACGTGATCTGAGCTTGAAGGGCACA 5040  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 DB 5041 GGGCTTCAAGGCTCTCTCTCTCGTGTGAGCGCAGGCGCGGCAATGACGATGATCGCAC 5100  
 QY 1701 HisAlaGlnLeuAlaLeuGlyValAsnSerGlyProGlyIysHisAlaIleLeuSerPheAsp 1720  
 DB 5101 CATGCAAGCTGGCACTGGAGGCCAGGGGGGCGCTGGCATGCAATTTGTCTTCGAT 5160  
 QY 1721 TYrGlyGlnGlnArgAlaGlyIuIysAsnLeuGlyProArgLeuHisGlyLeuHisIleuSer 1740  
 DB 5161 TATGGGCAAGCAAGAGAGAGGCAACTGGGCGCGGCTGCAATGTCTGCACTGAGC 5220  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 DB 5221 AACATTAACATGGGGGGAATACCTGGGCGAGCGCGGTGTGGCCCGTGTCTTGGGGGC 5280  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 DB 5281 TGTGTGAGGGTGTGCGGGTGAAGCAATACCGAGAGGGGTTTAAACAGCTGTGATCCAGC 5340  
 QY 1781 HisGlyIysSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB 5341 CATGGGAGAGCATCAAGTGGAGCAAGGCTGTATCTGTGACCTTGTGTATCAAC 5400  
 QY 1801 ProCysProAlaAsnSerTYrCysSerAsnAspThrAspSerTYrSerCysSerCysAsp 1820  
 DB 5401 CGGTGTCTGTAAAGCATATTTGACAGAACACATGGGACAGTATATCTTGGAGCTGTGAT 5460  
 QY 1821 ProGlyTYrTYrGlyAspAsnCysThrAsnValCysAspLeuAspProCysGlyIuIysGln 1840  
 DB 5461 CAGGTACTATGTGAGCAACTGTACTATGTGTGTACCTGAACCGCTGTGAGCACAG 5520  
 QY 1841 SerValCysThrArgIysProSerAlaProHisGlyTYrThrCysGluCysAspProAsn 1860  
 DB 5521 TCTGTGTGATCCCGAAGCCGAGTCCCGCCCATGGCTATACCTGTGAGAGTCCCAAT 5580  
 QY 1861 TYrLeuGlyProTYrCysGlnThrArgIleAspGlnProCysProArgGlyTYrThrGly 1880  
 DB 5581 TACCTTGGGCACTATCTGTGAGCCAGATGACAGCGCTTGTCCCGGTGGTGGGGA 5640  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerIysGlyPheAspProAspCysAsn 1900  
 DB 5641 CATCCACATGGGGCCCATGCACTGTATGTGACGAAGCAAGCTTGTGACCCAGACTGAC 5700  
 QY 1901 IysThrSerGlyGlyCysHisCysIysGluAsnHisTYrArgProProGlySerProThr 1920  
 DB 5701 AAGACAGAGGGGAGATGGCACTGCAAGAGAACTACAGCGGCCCCAGGCGAGCCCCACC 5760  
 QY 1921 CysLeuLeuCysAspCysTYrProThrGlySerLeuSerArgValCysAspProGluAsp 1940  
 DB 5761 TGCCTTGTGTGACTGTACCCCAAGAGCTCTTGTGCAAGTGTGTGACCTTAAGAT 5820  
 QY 1941 GlyIysCysProCysIysProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 DB 5821 GGCGAGTGTCAATGACAGAGTGTATCGGGCGCTCAGTGTGACCGCTGTGACCACT 5880  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTYrAspSerCysProArgAlaIle 1980  
 DB 5881 TTTGTGTAGGTACCAACCAATGGCTGTAAAGTAATTAAGACGTGGCCCAAGACGAT 5940  
 QY 1981 G1uAlaGlyIleTYrThrProArgThrArgPheGlyLeuProAlaAlaAspProCysPro 2000  
 DB 5941 GAGGCTGGATTTGGTGGCGCGCTTCCGCTTCCGAGCTGTGCTCTCTGTCTCC 6000  
 QY 2001 IysGlySerPheGlyThrAlaValArgHisCysAspArgIuHisArgGlyTYrLeuAspPro 2020  
 DB 6001 AAAGGCTCTTTGGAGACTGTGTGGCCCACTGTATGAGCAAGGGGGGTGTCTCCCA 6060  
 QY 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuIysGlyPheAlaGluArgLeu 2040

Db 6061 AACCTTTCAACGTCAGCTCCATCACCTTCCAGAACTGAAAGGCTTCGCTGAGCCGCTA 6120  
QY 2041 GlnArgAsnGlyuSerGlyLeuAspSerGlyArgSerGlnGlnLeuLeuLeuLeuArg 2060  
Db 6121 CAGCGAATATAGTCAGGCTTACCTGAGCTAGGAGGCTCCAGAGCTTACGCTGCTGCGC 6180  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIleValAlaTyrGlnLeu 2080  
Db 6181 AACGCCACCGACAGCACAGCTGGCTACTTGGCAGAGCACTCAAGGTGGCTTACAGCTG 6240  
QY 2081 AlaThrArgLeuLeuAlaHisGlyuSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6241 GCCACGCGCTGCTGGCCCAAGAGACCCAGCGGGCTTGGGCTGTCTGCAACAG 6300  
QY 2101 AspValHisPheThrGlnLeuLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6301 GAGCTGCACTTCACTGAGAACTGCTGCGGGTGGGAGCGCCCTCTGAGACAGCCAAAC 6360  
QY 2121 LysArgHisTrpGlyuLeuLeuLeuGlnGlnThrGlnGlyTyrAlaTrpLeuLeuGlnHis 2140  
Db 6361 AAGCGGCACCTGGAGCTGATCCAGACAGAGGGTGGCACTGGCTGGCTGCCAGCAC 6420  
QY 2141 TyrGlnAlaTyrTrpAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160  
Db 6421 TATGAGGCTTACCGCAGTGCCTGGCCAGAACATGCGGCACTTAACCCCTTC 6480  
QY 2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspIleGlyAsnPheAla 2180  
Db 6481 ACCATGTCACGCGCCAACTGTCACTCCGTAGTGCCTTGGACAAAGGAACTTGGCT 6540  
QY 2181 GlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
Db 6541 GGGGCCAAGCTGCCCGCTACAGAGCCCTGCGGGAGAGCGCCCGGACCTTGAGACA 6600  
QY 2201 ThrValIleLeuProGlyuSerValPheArgGlnThrProProValIleArgProAlaGly 2220  
Db 6601 ACAATATTCCTGCTAGTCTGTCTTCAAGAGACGCCCCGCTGTGACAGGCCCGGAGGC 6660  
QY 2221 ProGlyGlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgArgHisProGlnLeu 2240  
Db 6661 CCGGAGAGGCGCCAGAGGCGACAGAGGAGCTGGCACGCGACAGGACCGCGAGCTG 6720  
QY 2241 SerGlnGlyGlnAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db 6721 AGCGAAGGTGAGGCTGTGGCAGCTCATCTACCTGACCTGCGCGGCTTACGCTG 6780  
QY 2261 HisAsnTyrAspProAspIleArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
Db 6781 CATACCTATGACCTCTACAGAGCGACGCTTGAGAGTCCCAAGCCCGATCATCAACACA 6840  
QY 2281 ProValIleSerIleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspLys 2300  
Db 6841 CCGGTGTGAGCATCAGCGCTCATGATGAGAGACTTCTGCCCCGGGCCCTGAGCAAA 6900  
QY 2301 ProValIleValGlnPheArgLeuLeuGlnThrGlnGlnArgThrLysProIleCysVal 2320  
Db 6901 CCGGTACCGGTGACGTTCCGCTGAGAGACAGAGGAGGACCAAGCCCATCTGTGTC 6960  
QY 2321 PheTrpAsnHisSerIleLeuValIleSerGlyThrGlyGlyTyrSerAlaArgGlyCysVal 2340  
Db 6961 TTCTGGAAACCATTCATCCCTGTCAGTGGACAGGTGGCTGTGCGCCAGAGGCTGTGAA 7020  
QY 2341 ValIlePheArgAsnGlyuSerHisValIleSerCysGlnCysAsnHisMetThrSerPheAla 2360  
Db 7021 GTCTGTCTTCGGAATAGAGCCAGCTCAAGTGCACAGTGCACCATGACAGACTTCGCT 7080  
QY 2361 ValLeuMetAspValIleSerArgArgGlnAsnGlyGlnIleLeuProLeuLysThrLeuThr 2380  
Db 7081 GGCTCTATGAGACCTTCTCGCGGAGAAATGGGGAAATCTGCTGACCTGAAAGCACCTGAA 7140  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400

Db 7141 TACGTGGCTTAGAGTGTACCTTGGCTGCCCTTGTGCTACCTTCTTCTCACTCTTC 7200  
QY 2401 LeuArgIleLeuAspSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleuGly 2420  
Db 7201 TTGGCTATCTCGCTCCCAACCAACGAGCATCCAGATCTGACCTGACCTGCTGGGC 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
Db 7261 CTGGCTCAGCTGATCTTCTCTCGGGAAATCAACAGAGGCTGACTCCCTTTGTGCTGACA 7320  
QY 2441 ValIleIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGln 2460  
Db 7321 GTCATTTGCCATCTGTGACCTTCTGATCACTTGTACCTTCTTCTGGGCTGTGCGAG 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValIleAspAspValAsnThrGlyProMetArg 2480  
Db 7381 GCTTGTACCTTACCGGCACTCACTAGAGTGGCGGATGTACACCGGCCCCATGCCC 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db 7441 TTCTACTACATGCTGGGCTGGGGCGTGCCTTCATCAGAGGCTAGCCGTGGGCTG 7500  
QY 2501 AspProGlnGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520  
Db 7501 GACCCGAGGGCTACGGGAACTGTGACTGTGGCTGTCTCATCTATGACAGCTCATC 7560  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
Db 7561 TGGAGTTTTCGTGGCCCGGTGGCTTTGGCGCTGTGATGAGTCTTCTGTACATCTG 7620  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyuLysGlyProValSer 2560  
Db 7621 GCGGCCCGGCTCTCTGTGTGCTGCCAGCGGACGGCTTGAAGAAAGGCTGTGCTG 7680  
QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
Db 7681 GGCTGTGACGCTCTCTGCGCTCTGCTGTGAGCGCACGTGGCTGTGGACATG 7740  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
Db 7741 CTCTGTGACAGAGGACACCTCTCTTCACTCATCTTGTGCACTGCAATGGCATC 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerLysGlnValArgLysAlaLeu 2620  
Db 7801 CAGGCGCCCTTCATCTTCTCTCTATGTGGCTTAGACAGAGGTCCGGAACACATC 7860  
QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
Db 7861 AAGCTTGGCTGAGCGGACGAGCCCTGACCTGTGCTGACCAACCAAGTCAACCTG 7920  
QY 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db 7921 ACCTGTCTTCAACTGTGCCCAAGCCCTACGAGATGGGCGGCTTACAGCCCTTACGGA 7980  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
Db 7981 GACTGCGCGGCTCTCTGACAGCACAGTCCCTGCGGCAAGATCAGGCCAGTACATC 8040  
QY 2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db 8041 CCGTCTTGTGAGGAGAGAGTCCGACCTGAAACCTGTGCAAGGGGCCCTGTGGCTGGGG 8100  
QY 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8101 GATCCAGGACGCTGTCTCTGGAAGTCAAGACAGCATGATCTGACACGGAATCC 8160  
QY 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
Db 8161 GACAGTGAACCTGTCTTGAAGACCAACAGAGTGTCTTATGCTTACCACTCATCA 8220  
QY 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
Db 8221 GACAGTGAAGAGAAAG 8280

QY 2761 TTPApSerLeuLeuGlyProGlyValAglyuArgLeuProLeuHisSerThrProLyAsp 2780  
 Db 8281 TGGATATGCTGCTGGGGCCCTGGAGCAGAGAGACTGGCCCTGCACAGTACTCCCAAGAT 8340  
 QY 2781 GlyGlyProGlyProGlyValAspAlaProThrProGlyAspPheGlyThrThrAlaLeuGly 2800  
 Db 8341 GGGGGCCCAAGGGCCCTGGCAGAGGGCCCTGGCCAGAGAGCTTGGGAGCCACAGCAAAAGAG 8400  
 QY 2801 SerSerGlyAsnGlyAlaProGlyGluLysArgLeuArgGlyLysAsnGlyAlaLeuSerArg 2820  
 Db 8401 AGTATGGCAACGGAGGCCCTGAGAGAGGCTGGGAGAGATGAGATGCTGTCTCGA 8460  
 QY 2821 GluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisGlyGlyLeuLeuLys 2840  
 Db 8461 GAGGGGTCCCTAAGGCCCTTCCAGGGCTCTTCCGCCAGCTCCACAAAGGATCTTAAAG 8520  
 QY 2841 LysLysCysLeuProThrThrLysSerGlyLysSerSerLeuLeuArgLeuProLeuGln 2860  
 Db 8521 AAGAAAGTCTGCTCCACCATCAGCAGAGAAAGCAGCCCTCGCGCTCCCTGAGACAA 8580  
 QY 2861 CysThrGlySerSerArgGlySerSerAlaSerGlyGlySerArgGlyGlyProProPro 2880  
 Db 8581 TGCACAGGATCTTCCGGGGCTCTCCGCTAGTGAAGGAGAGCCGGGGGGCTCCCTCC 8640  
 QY 2881 ArgProProArgGlnSerLeuGlnGlnLeuAsnGlyValMetProLeuAlaMet 2900  
 Db 8641 CGCCACCGGCCCGGAGAGCTCCAGAGAGCAGTGAACGGGGTCAATGCCATGCCATG 8700  
 QY 2901 SerLeuValAlaGlyThrValAspGluAspSerSerGlySerGlyPheLeuPheAsn 2920  
 Db 8701 AGCATCAAGGACGAGCAGGTGATGAGAGACTCGTCAAGCTCCGAAATTTCTTCTTAAC 8760  
 QY 2921 PheLeuHis 2923  
 Db 8761 TTCTCTGCAAT 8769  
 RESULT 2  
 AX549238 10531 bp DNA Linear PAT 26-NOV-2002  
 LOCUS Sequence 523 from Patent WO02061087.  
 DEFINITION AX549238  
 ACCESSION AX549238  
 VERSION AX549238.1 GI:25813930  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHOES Burmer, G.C., Roush, C.T., and Brown, J.P.  
 TITLE Antigenic peptides, such as for G protein-coupled receptors  
 (GPCRs), antibodies thereto, and systems for identifying such  
 antigenic peptides  
 JOURNAL Patent: WO 02061087-A 523 08-AUG-2002;  
 Lifespan BioSciences, Inc. (US)  
 FEATURES  
 source 1..10531  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 2038 a 3353 c 2998 g 2142 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 10531  
 Score: 15545.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-916-849a-3 (1-2923) \* AX549238 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 Db 63 ATGGAGAGCCCGGCCACCGGGCTCTCCACAGCCCGCGCGCTGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 123 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSer 60  
 Db 183 GGGTCCAGGGAGAGAGCTCTTGGGGGCTGGGCCCCCAATGGGCTGCTGTCCATTC 242  
 QY 61 SerAlaSerAsnLeuThrLeuThrLeuThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 243 TCAGGCTTCGAACTCTGGCTTACACACACCGCTGAGAGGATGGGCACTGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuAsnGlyValTrpCysProGluSerGlyAlaHis 100  
 Db 303 GGGCACCTGGTACCCCAACAGATGCTGAGAGGTTGGGTTCAGAAATCCGAGGCCAT 362  
 QY 101 IleProLeuProProAlaProGluGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
 Db 363 ATTCCTCTACACAGCTCTCTGAGAGCTCCCTGAGCTGTGCTCTGCGCATTTGA 422  
 QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGluGlnHisAspProCysLeuValAla 140  
 Db 423 GGGCACCTTCCCAACAGGGCAAGCTCACATGCCGAGAGAGACCCGTTAAAGGCT 482  
 QY 141 ProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLysLeuArgAlaGlyGlu 160  
 Db 483 CCACGGCTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542  
 QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgLysArgAsnValAsnThrAlaProGln 180  
 Db 543 AGGTCAACAGAAAGTCTCTGGGTGGGCTCCGAAAGAAATGTAAATCAAGCCCCCAG 602  
 QY 181 PheGlnProProSerTyrglnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db 603 TTCCAGCCCCCAGCTACAGAGCCACAGTCCGGAGAACACAGCAGCAGCAGCAGCAG 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGlyGlyValAlaGlyArgLeuGlnTyrrMet 220  
 Db 663 GCATCCCTTGAAGGCCATTCACCCGAGCAGAGGTGAGGAGGTGACTGAGTACACAG 722  
 QY 221 AspAlaLeuPheAspSerArgSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 723 GATGCCCTTTTATAGCCGCTCCACACAGTCTTCTCCCTGAGACCACTACTGTGCA 782  
 QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrLysSerThrHisValPheArgValThr 260  
 Db 783 GTACACACAGCCAGAGAGCTGATGCTGAGACCAAGAGCACCACGCTTCAAGGGTCAG 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaAlaValAlaThrLeuThrIleLeuValThr 280  
 Db 843 GCGCAGAGCACAGCAGATGCCCGCAGAGAGTCTCGGTCACTCAACATCTTGTTACT 902  
 QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlnTyrrLysGlySerLeuArgGlu 300  
 Db 903 GACACCAAGACACATACCTCTGTTCAGACGACGAGTACAGAGAGGCTCAGAGAG 962  
 QY 301 AsnLeuGlnValGlyTyrglyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 963 AACCTGAGAGTGGCTATAGGTGCTCACTGTCAAGGGCCACAGATGTATATCCCTCCC 1022  
 QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGlyLysSerGlySerProSerGlyValPhe 340  
 Db 1023 AATGCCAATATCTGATCCGCTGCTGAGAGGGCTCTGGGGGAGGCCCTTCAAGTCTTT 1082  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlyGluVal 360  
 Db 1083 GAGATGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142  
 QY 361 GluSerTyrglnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380

```

Db      1143 GAATCTTACCAAGCTGAGGAGGAGGACCAAGGCTCGGAGCCCGGCTCTCGAAGT 1202
QY      381  ThrThrAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400
Db      1203 ACCACAGCGCGCTGTTTCTTCTTGAGGAGATGACATATATATGCCCCCGATTAAGT 1262
QY      401  GluLysArgTyrValValGlnValArgGluAspValThrProGluValProValLeuArg 420
Db      1263 GAAAGACGCTATGTGTCCAGGTGAGGAGATGATCTCAGGGCCCCAGATACCCGA 1322
QY      421  ValThrAlaSerAspArgAspLysArgLysSerAspAlaValIleIleTyrSerIleMetSer 440
Db      1323 GTCAAGCCTCGGATGAGACAAAGGAGCAATCGGTGTGACATATGACATCAAGAGT 1382
QY      441  GlyAspAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460
Db      1383 GGCAGAGCTCGGGGACAGTTTATCTGAGATGCCCAAGCTGGACCTTGAGATGCGTGAAC 1442
QY      461  ProLeuAspTyrGluThrThrTyrGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db      1443 CCTCTTGACTATGACGACCAAGAGATACCCCTACGGGTGCGAGCAAGATGGTGGC 1502
QY      481  ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleLeuAsp 500
Db      1503 CGTCCCACTCTTAATGTCTGTGCTGTGATGACAGTACAGGTCTCTGATATCAACGAC 1562
QY      501  AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      1563 AATGGCCCATCTTCTGTCAGCAACCCCTTTCAGAGCTACTGTCTGAGAGAGCGCCCTTA 1622
QY      521  GlyTyrLeuValLeuIleValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db      1623 GGGTACTGCTGTTCTCATGTCCAGGCTATCGACGTGATGCTGTGATCAATGCGCCCTG 1682
QY      541  GluTyrArgLeuAlaGlyValGlyIleAspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1683 GAATACCGCTTGTGAGGAGGAGCAATGCTCCCTTCAACATCAACATGACACAGGC 1742
QY      561  TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      1743 TGGATCTCTGTGCTGCTGATCGACCGGAGAGAGTTGATTTCTACAGCTTTGGGAT 1802
QY      581  GluAlaArgAspIleGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      1803 GAGAGCTCGAGACCATGACCTCCAGCACTCACTCGCCGCGCAGTCTCAGCGTCACTGTC 1862
QY      601  LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      1863 CTGGATGTCAACGACAAACATCCACCTTTACCCAAACAGAGTACACAGTGCAGCTCAAT 1922
QY      621  GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      1923 GAGGATGACGCTGTGGGACACAGCGTGTGACGCTGTCAGCTGTGACCGTATGCTCAT 1982
QY      641  SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1983 AGGTGATCATCACTTACAGATCACACAGTGGCAATATCGAAACCGCTTCTCATCAACAC 2042
QY      661  GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnValArgGln 680
Db      2043 CAAGAGGTGTGTGGCTGGATCCCTTGCCCTGACCTGACATCAAACTTACGCGGACG 2102
QY      681  TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      2103 TATGTGTGGCTGTATCCGCTCGATGGCACTGGCAGGACACCGCACAGATGTGGTG 2162
QY      701  AsnValThrAspAlaAsnThrIleArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      2163 AATGTACCGACCGCAACCCATCGTCTCTTCAAGGCTCCCACTATTCAGTGAAT 2222
QY      721  ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740

```

---

```

Db      2223 GTTAATGAGACCGGCGGACGACCAACGCTGTGCTGATACGGCCACGATGAGAC 2282
QY      741  ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2283 ACGGTGAGAAAGCCCGCATCCATCTTATGAGAGACAGCATCCCGAGTTCCGCATC 2342
QY      761  AspAlaAspThrGlyValAlaThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2343 GATGACAGACCGGGGCTGTCAACCAAGGCTGAGTGTGATTAAGAGACCAAGTGTCT 2402
QY      781  TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleLysSerAspThrThrTyr 800
Db      2403 TACACCTGGCCATTATCTGCTCGGAGCAATGGCATTTCCAGAAATCGGACCAACCTAC 2462
QY      801  LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db      2463 CTGAGATCTCTGTGAAACGATGAAATGACATGCCCCCTCACTCTCTGAGATCTCTAC 2522
QY      821  GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2523 CAGGCGACGTCTATGAGAGATGTGCCACCTTCACTACGCTCTGACATCTCAACCACT 2582
QY      841  AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db      2583 GATCGTGAATTCGACCTTAATGGCAGGGTCTTCAACCTTCCAGAGGCGACATGGA 2642
QY      861  AspGlyAspPheIleValGlySerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db      2643 GACGCTGACTTATTTATGTTGATGCCACGTCAAGCATCGTGCAGACGCTACGAGGCTGAT 2702
QY      881  ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900
Db      2703 CGAGAGAACGTGGCCCAATATGTCTTGGGCAATGATGACAGAGGAGATGCCCCCA 2762
QY      901  AlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db      2763 GCCCGCAACCTATGAAAGTACAGTCACTGTGTGATGTGAATGACAAATCCCTCTGTC 2822
QY      921  PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2823 TTTGAGCAGAGATGATTTGATGTGTTTGTGAGAGAGAACACCCCATTTGGGTACCGTGT 2882
QY      941  AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      2883 GCCCGGCTCAAGCCACGACCCCGATGAAAGCCCAATGCGCATATATGACACAGATT 2942
QY      961  ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db      2943 GTGAGGGCAACATCTCGAGGCTCTTCCAGCTGACATCTTCTCCGGGAGCTGACAGCC 3002
QY      981  LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db      3003 CTGGTGAACCTTAGACTAGAGAGACCGGCTGAGTACGCTCTGTGATCAAGGCCACGTTCA 3062
QY      1001  AlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAsnAspAsnPro 1020
Db      3063 GCTCTCTGTGAGAGCGGCTTACAGTCCAGCTCGCCCTCTTGAACCGCAATGACAAACCA 3122
QY      1021  ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db      3123 CCAAGTCTGGGCACTTGAATCTTTTCAACAACATATGTCAACCAATGCTCAAGACAGC 3182
QY      1041  PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3183 TTCCCTGGGGGGGCGCATTTGGCGAGTACCTGCCATATGACCTCGATATCTCAGATGCTGTG 3242
QY      1061  ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      3243 ACTTACAGCTTGTAGCGGGGAAATGAACTCAGCTGTGCTCCTCAATAGCTCCACCGGGT 3302
QY      1081  GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db      3303 GAGCTGAAGCTTAGCGCGGCACTGACACAAACCGGCTCTGAGAGCCATCATAGCGTG 3362

```



QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
 Db 3363 CTGGGTGACAGCGGCTACACAGCGTGACCGGCCGCGCTCGGTGATCACATCATC 3422  
 QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
 Db 3423 ACCGATGAGATGCTCACACGACGATCACGCGCCCTGAGAGACATGTCTACCCGAGCGC 3482  
 QY 1141 PheLeuSerProLeuLeuGluValPhePheIleGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
 Db 3483 TTCCGTGACCACTGCTAGGCTCTTCAATCCAGCGGTGGCCGCCACGCTGACCGCA 3542  
 QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGluGlyHisIle 1180  
 Db 3543 CCGGACCAAGTGTGTCTTCAACGTACAGCGGACACCGACGCCCGGGGGCCACATC 3602  
 QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGluProGluGlyGlyProProPheLeu 1200  
 Db 3603 CTCAACGTGACCTGTGCGTGGCCAGCGCCGCGGGCGGGCGCGCCCTTCTCTG 3662  
 QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
 Db 3663 CCTCTGAGACCTGTCAGAGCGCTTATCTTCAACCGACGCTGTCAGCGCCATCTCG 3722  
 QY 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240  
 Db 3723 GCAACAGCGCGTGTGCTCTTCCGACGACAACTGCTCGCGGAGCCCTGAGAACTAC 3782  
 QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
 Db 3783 ATGGCGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3842  
 QY 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGluPhe 1280  
 Db 3843 GTGCTCTTCCGCGCATCACCCGCTCGGAGGGCTGGCTGCGCTGCGCCCGCGCTTC 3902  
 QY 1281 ThrGluAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3903 ACGGCTGACTACTGAGAACCGAGGTGAGCTGTGCTACTCGCGGCGCTGTGGCCCCAC 3962  
 QY 1301 GlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 Db 3963 GGGGCGCTGCGGACCGCGAGGCGGCTTACACTGCTCTGTGTGTGTGTGTGTGTGTGT 4022  
 QY 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGluValCysValAsnGly 1340  
 Db 4023 GAGCACTGTAGTGATGCTGCTCGCTCAGGCCGCTTGCACCCGGGTGTGTGAAATGG 4082  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACCTGTGTCAACTGCTGTGGCGGTTCAAGTGCATTCGCCCATCTGAGACTTC 4142  
 QY 1361 GluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 Db 4143 GAGAAAGCCCTACTGTCAGATACACAGCGAGCTTCCCGCCACTCTTATACACCTTT 4202  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGluArg 1400  
 Db 4203 CGCGGCGTGGCGGCGTTTCACTTCACTTGCCTCTCTGTTGTCACAAAGAGCGC 4262  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyValArgPheAsnGluLysHisAspPheValAlaLeuGlu 1420  
 Db 4263 GACCGGT 4322  
 QY 1421 ValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440  
 Db 4323 GTGATCCAGAGACAGGTCAAGCTCACTTCTGTGAGGAGAGTCAACACACAGCTGTCC 4382  
 QY 1441 ProPheValProGluGlyValSerAspGlyGlnThrHisThrValGlnLeuLysTyrTyr 1460  
 Db 4383 CCAATTCGTCGCCGAGAGAGT 4442

QY 1461 AsnLysProLeuLeuGlnGlnThrGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
 Db 4443 AATAGCCACTTGTGGTTCAGACAGGCTTCCACAGGCCCAACAGAGAAAGTGTGT 4502  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 Db 4503 GTGTGTGACCGTGTGATGT 4562  
 QY 1501 AsnTyrSerValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1520  
 Db 4563 AACTACTCTGT 4622  
 QY 1521 GlyProLeuLeuLeuGlnGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
 Db 4623 GGGCGCTGTCTTAAAGCGGGGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4682  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 Db 4683 CAGTTCGTGGCTGT 4742  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
 Db 4743 TTCAATGCCAAATGTGACAGT 4802  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
 Db 4803 AACACTTGCCACAATGGGGGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4862  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
 Db 4863 CCCCCTGGCTTTGGGGGCAAGACTGCGCCAGAAATGTCCAAATCACACGCACTTCTG 4922  
 QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640  
 Db 4923 GCGACAGCTGT 4982  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 4983 AGCCTCATGTTCGCGACGCGCCAGCGCGGAGGTGTCTGTGTGTGTGTGTGTGTGTGT 5042  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlnGlnHisValMetLeuSerValGlnGlyThr 1680  
 Db 5043 CGGACCACTATCCCTTACGACTACAGAGGGCGACGTATGTGTGAGCGTGGAGGACCA 5102  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGluArgAlaAsnAspGlyAspThrHis 1700  
 Db 5103 GGGCTTACAGGCT 5162  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 5163 CATGCAAGCTGTGCACTGGGAGCGCACGCGGGGGCTGTGCCATTCGTGTCTTCGAT 5222  
 QY 1721 TyrGlyGlnGlnArgAlaGlnGluAsnLeuGlnProArgLeuHisGlyLeuHisLeuSer 1740  
 Db 5223 TATGGGAGAGAGACAGAGAGGCAACTGTGGCCCCCGGTGTGTGTGTGTGTGTGTGTGT 5282  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
 Db 5283 AACATAACAGTGGCGGGAATCTGTGGCCAGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 5342  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 Db 5343 TGTTCAGAGGT 5402  
 QY 1781 HisGlyGlnSerIleAsnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1800  
 Db 5403 CATGGGAGAGCACTCAAGTGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5462  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAsp 1820  
 Db 5463 CCGTGTCTGT 5522  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840



```
Db 5523 CAGGTTACTTATGATGACACTGTAATGATGACCTTAACCCGGTGAACAACG 5582
Qy 1841 SerValCysThrArgIleuPheProSerAlaProHisGlyTyrThrCysGluCysProProAla 1860
Db 5583 TCTGTGTGATACCGGAGGCGCAAGTGGCCCCCAATGGCTAATACCTGGGAGATGTCCTCCCAAT 5642
Qy 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrTrpGly 1880
Db 5643 TACCTTGGGCACTTACTGTAGACCAAGATTGACCAAGCTTGTCCCGTGGCTGGTGGGA 5702
Qy 1881 HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn 1900
Db 5703 CATCCCAATGTGGCCCATGCAACTGTATATGTCAAGAAAGCTTTGACCCCAACTGCAAC 5762
Qy 1901 TyrThrSerGlyGluCysHisCysGlyGluAsnHisTyrArgProProGlySerProThr 1920
Db 5763 AAGACAAGGGGAGTGGCACTGCAAGAGAACCACTACCGGCCCCCAAGGACCCCAAC 5822
Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940
Db 5823 TCCCTCTGTGTGACTGCTAACCCCAAGGCTCTTGTCCAGAGTCTGTGACCTTGAGAT 5882
Qy 1941 GlyGlnCysProCysIleuProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960
Db 5883 GGCACAGTGTCCATGCAAGCAAGGTGTCATCGGAGCGTCACTGTGACCGTGTGAACAACCT 5942
Qy 1961 PheAlaGluValThrThrAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaIle 1980
Db 5943 TTTGTGAGGTACCAACCAATGGCTGTGAAGTAATGACAGCTGGCCCAAGAGCGATT 6002
Qy 1981 GluAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000
Db 6003 GAGGCTGGGATCTGGTGGCCCGTACCCGCTTCGGGCTGCTGGTGGCTCCCTGTC 6062
Qy 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArgGlyTyrLeuProPro 2020
Db 6063 AAGGCTCTTTGGGACGTGCTGGCCACTGTATATGACAGAGGGGAGTGCCTCCCA 6122
Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuGlyPheAlaGluArgLeu 2040
Db 6123 AACCTTTCACATGCAACGTCATCACTCTTCAGAACTGAAGGCTTGTGTAAGGCGGTA 6182
Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060
Db 6183 CAGCGGAATGAGTCAAGGCTTACAGTCAAGGCGCTCCCAAGCACTGCTCTCTGGGCG 6242
Qy 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080
Db 6243 AACGCAAGCAAGCAAGCTGCTACTTGGCAGCGAGCTGAAGGTGGCTTACCAAGCTG 6302
Qy 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db 6303 GCCAAGCGGCTGCTGGCCCAAGAGAACCCCAAGCGGGCTTTGGGCTGTCTCCACAGAC 6362
Qy 2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db 6363 GACGGTCACTTCACTGAGATCTGCTGGCGGTGGCAGGCGCTCTGGACACACCAAC 6422
Qy 2121 LysAspGlnIleTyrGlnLeuIleGlnGlnThrArgIleGlyTyrAlaIleThrLeuGlnHis 2140
Db 6423 AAGCGCACTGGGAGCTGATCCAGAGAGAGAGGGTGGCAGCGCTGGCTGTCCAGAC 6482
Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160
Db 6483 TATGAGGCTTACGCAAGTGCCTTGGCCCAAGAACAGCGGCACTTACTTAAGCCCTTTC 6542
Qy 2161 ThrIleValThrProAsnIleValIleSerValValAlaGluAspLysGlyAspPheAla 2180
Db 6543 ACCATCGTCAACCGCAACATGTTCATCTCCGTATGGCTGTGGACAAAGGGAATCTTGTCT 6602
Qy 2181 GlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlyGlnProProAspLeuGlnThr 2200
Db 6603 GGGGCAAGCTGCCCGCTTACAGGCGCTGCGTGGGAGACAGCCCCCGGACCTTGAGACA 6662
Qy 2201 ThrValIleLeuProGluSerValPheArgGlyThrProValValArgProAlaGly 2220
Db 6663 AACATGATTTCTGCGTGAATCTGTCTTCAAGAGAGCGCCCCCGTGTGACAGCGCGCAAGC 6722
Qy 2221 ProGlyGluAlaGlnGlnProGluGluLeuAlaArgArgGlnArgArgHisProGluLeu 2240
Db 6723 CCGGAGAGGCGCCAGAGACCAAGAGAGCTGGCACGCGCAAGCGACCGGAGGCTG 6782
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260
Db 6783 ACGCAGGTGAGCTGTGGCCAGCTCATCATCTTACCGCACCTTGGCGGCTACTGCTCT 6842
Qy 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280
Db 6843 CATACATATGACCTTGACCAAGCGAGCTTGAGAGTCCCCAAGCCGATCATCAACACA 6902
Qy 2281 ProValValSerIleSerValHisAspAspGluLeuLeuProArgAlaLeuAspLys 2300
Db 6903 CCGGTGTGACATCAGCGTCCATGATGATGAGAGCTTCTGCCCCGGGCGCTGACAA 6962
Qy 2301 ProValThrValGlnPheArgLeuGlnThrGlnGlyArgThrLysProIleCysVal 2320
Db 6963 CCGGTCAAGTGTGCTGCTGCTGTGAGACAGAGAGCGGACCAAGCCATCTGTCTC 7022
Qy 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGlu 2340
Db 7023 TTTGTGAACATTCATCTGTGTGATGAGTGCACAGTGGCTGTGGCCAGAGGCTGTAA 7082
Qy 2341 ValValPheArgAsnGlySerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360
Db 7083 GTGCTTCCCAATGAGAGGCAAGTCAAGTGCAGGTCGCAACATGACAGGCTTCCCT 7142
Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGlyIleLeuProLeuLysThrLeuThr 2380
Db 7143 GTGCTATGACGATTTCTCGCGGAGATGGGAGATCTGGCCACTGAACACTGACA 7202
Qy 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400
Db 7203 TACGTGGCTAAGGATGACACTTGGCTGCTGTGCTGCTTCTTCTCTCACTCTC 7262
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420
Db 7263 TTGGTATCCGCGCTCCAAACACAGGCACTCGACATGACAGCTGCGCTGGGCG 7322
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db 7323 CTGGCTAGCTGTGCTCTCTGTGAATCAACAGGCTGACCTCTTGTGGCTGACACA 7382
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460
Db 7383 GTCATTTGCATCCGTGCACTTCTGTACTCTGCACTTTTCTGGGCTGTCTGGAG 7442
Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480
Db 7443 GCTTGGACCTGTACCGGCACTCACTGAGGTGGCAATGTCACACCGGCCCCATGGCG 7502
Qy 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db 7503 TTCTACTCATGCTGTGGCTGGGCGTGGCTGCTTCAACAGGAGCTTGAAGCGGCTG 7562
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520
Db 7563 GACCCCGAGGGCTACGGGAACCTTGACTTGTGCTGCTTCCATCTATGACAGCTCATC 7622
Qy 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db 7623 TGAATTTGTGTGGCCCGGTGGCTTTGGCTTCAGATGATGTCTTCCGTGTACATCTG 7682
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnIlePheGlnLysGlyProValSer 2560
Db 7683 GCGGCGCGGCTCTCTGTGCTGCCACGCGGCGGCTTTTGAAGAAAGTCTGTCTGCG 7742
```

QY	256	GIYLeuGIInPProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu	2580
Db	7743	GGCCTGACGCCCTCTTCTGCGCGTCTCTGCTGCTGACGGCAGCGTGTCTGTGCACTG	7802
QY	2581	LeuSerValAenSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAenCysIle	2600
Db	7803	CTCTCTGTCAACAGACGACACCTCTCTTCCACATCACTTTGTGCTACGTGAATTGACATC	7865
QY	2601	GIInGlyProPheIlePheLeuSerTyrValValLeuSerTyrGluValArgValAlaLeu	2620
Db	7863	CAGGACCCCTTCATCTTCTCTCTATGTGTGCTTACACAGAGTCCGGAAGAAGCATCTC	7922
QY	2621	LYLeuAlaCysSerArgIysProSerProAspProAlaLeuThrThyLeuSerThreLeu	2640
Db	7923	AAGCTTGCTGCAACCGCAGAACCCAGCCCTACCTGTCTGACCAACCAATCCACCTG	7983
QY	2641	ThreSerSerTyrAenCysProSerProTyrAlaAspGlyValGlyLeuTyrGlnProTyrGly	2660
Db	7983	ACCTCGTCTTCAACAATGCCCCAGCCCTTACCCAGATGGCGCGCTGTACACACCCCTCA	8042
QY	2661	AspSerAlaGlySerLeuHisSerThreSerArgSerGlyLysSerGlnProSerTyrIle	2680
Db	8043	GACTCGCCGGCTCTCTGCAACAGACCACTGCTCGGCAAGATGACACCCAGCTACATC	8102
QY	2681	ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly	2700
Db	8103	CCCTTCTTGCTGAGGAGAGATCGGCACTGAACCTGTGCAAGGCCCCCTGGCCTGGG	8162
QY	2701	AspProGlySerLeuPheLeuGluGlyGlnAspGlnHisAspProAspThrAspSer	2720
Db	8163	GATCCAGCGACACCTGTGTTCTCTGAAGGTCCMAACCAACGACGATGATCTTCAACGCACTCC	8222
QY	2721	AspSerAspLeuSerLeuGluAlaAspAspIAsnSerGlySerTyrAlaSerThHisSerSer	2740
Db	8223	GACAGTACCTGTCTTGAAGAAGCAACAGATGTGCTCTTATGCTTCAACCACTCAATCA	8283
QY	2741	AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGly	2760
Db	8283	GACAGTACGAGGAGAAAGC	8342
QY	2761	TrpAspSerLeuLeuGluGlyProGlyValAlaGluTrpLeuProLeuHisSerThyProLysAsp	2780
Db	8343	TGGATATAGCTCTCTGGGGCTCTGAGACAGAGACCTGCCCTTGACACATCTCCAGAGAT	8402
QY	2781	GlyGlyProGlyProGlyLysAlaAspTrpProGlyAspPheGlyThrAlaIysGlu	2800
Db	8403	GGGGGCCACAGGGCTGTGGCAAGGCCCTTGCCAGAGCATTTTGGGACCAACGCAAAAGAG	8462
QY	2801	SerSerGlyAsnGlyValAspProGluGluArgLeuArgLeuAsnGlyAspAlaLeuSerArg	2820
Db	8463	AGTATGTGCACACGGAGCCCTCTGAGAGAGGGCTGCGGAGAAATGAGAGAGCCCTGTCTGCA	8522
QY	2821	GluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisIleGlyIleLeuLys	2840
Db	8523	GAGGGGTCTCCATGAGCCCTTCTCAAGGCTCTTCTGCCAGCTCAACAAAGGATCTTTAAG	8583
QY	2841	LysIysCysLeuProThrIleSerGlyLysSerSerLeuLeuArgLeuProLeuGluGln	2860
Db	8583	AAGAAGTGTCTGCCACCATCAAGCAGAGAGAGACGCTCTCGGGCTCCCCCTGGAGCAA	8642
QY	2861	CysThrGlySerSerArgGlySerSerAlaSerGluGlyLysSerArgGlyGlyProProPro	2880
Db	8643	TGCACAGGCTTCTCCCGGGGCTCTCTCCCTATGTAGGGCAGCGGGGGGGGCCCTTCCC	8702
QY	2881	ArgProProAlaGlnSerLeuGlnGluLeuAsnGlyValMetProIleAlaMet	2900
Db	8703	CGCCCAACGCCCCGGCAGAGCTTCCAGAGACGCTGAACGGGGATCATGCCATGCGCATG	8762
QY	2901	SerIleLysAlaGlyThrValAspGluAspSerSerGlySerGluPheLeuPheAsn	2920
Db	8763	AGCATCAAGGAGGACGATGATAGGATCTGTCAAGGCTCCGAATTTCTCTTCTTTAAC	8822

[illegible]

SRHIDMADPFIANNGTVCPCPAKKNVCDSTNCHNGTCVNDMDAFSCBCEPLFGKSCA  
QEMANPDRHSSLVAMHGLSLPISQPMVLSIMFRTQADGVLLQAIIRGSTTLOL  
REGVMHLSVEGTGLQASSLRLEPGANGDMHMOALAGSGPGHAILISDYGGORA  
EKNLGPRLHGLHLSNITVGGIIPGPAVGPRGCTGOCVRVSDPBGNNLSPSGES  
INVEGCSLPDPCDSNPPCANSCNDMDYSICSDPRTYGDNCTNVCIDLNPCHQSV  
CTRKSAPHGTTCECPNITLGPCEIRIDQPCRGMMWHPTGCPNCNCSVSGFDDCN  
KTSQCHCKENHRRPSPPTCLCDQCYTGLSRVCDDEDCQPCPKGVIRGOCDCD  
NPFAYVINGEVNVDSCPRAIENGIWMPRTFGLPAAPCPKGVTAIVHCBHRG  
WLPVLPNCTSIITPSLKGPAERLQNRNSGLDSGQOALLRNTAOTHTGYPGSDV  
KVAVOLATRLAHSTORPGLSATODHPENLRLVGSALLDPTAKHMLI00TEG  
GTAVILOHYEAVASALAQMRHTYLSPTTYTPNVIVSVVLDKENPRGACLPRTBAL  
RGEOPDLETTYILPESVFRTPVVRPVRKGEARQERBELH0QGEVAVS  
VI1KRTLAGLPHNDVPRKSLRVPKRIINPVVSIHDEBELPALRKYVVOF  
RLBETERTKPICVFMNHSILVSGWSARGCEVFNESVSCQCHMTSPAVLMD  
VSRNENGILPLKTLTVALLGTALALITFFFLRLILRNSNOHIGIRNTALAGLA  
OLVFLGINODLPACTVIALILHRLICTPSWALLALILYVALTEVRVNTGPMR  
FYVTLGKGVPAFTINGLAVGIDBEGNDPFCMLSTYDPLINSFAPVAPANMSVFLY  
ILABASCAADQREQKGVPSGLQPSFAVILLISATWLLLSVNSDTLLPHYLPAT  
CNCIQGPPIPLSYVLSKEVRKALKLACSRKSPDPAITTSITLSYNCSPPADQR  
LYOPYGDAGSLHSTRSKQSPSYIPLRLERSALNFGQPGIAGDPSLFLRGDQ  
QHDPTDSDSLSDQSGYASTHSDESEEBEBAAPFGQCMDSLILGGA  
RLPLHSTKDGSGPGKAPWPGDPTGTAKESNGAPERRLRNENDLSREGSGLPL  
GSSAQPHKGIKKKCLPTISKSSSLRLPLEGCTSSSRGSSASBSRGPPPPPPRO  
SLQBLNVMPIHMSIKAGIVDESDSSSEFLFPNFLH"

BASE COUNT 2038 a 3353 c 2998 g 2142 t

## ALIGNMENT SCORES:

Pred. No.: 0 Length: 10531  
Score: 15545.00 Matches: 2923  
Percent Similarity: 100.00% Conserved: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AF234887 (1-10531)

QY 1 MetArserProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
DB 63 ATGGCGAGCCCGGCGACCGGGGCTCCCTCCCAACCGCGCGCGCTGTGTG 122  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyValGlyProGlySerLeu 40  
DB 123 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyTrpLeuCysProSer 60  
DB 183 GGGTCACGGGACGAGGCTCTTGGGGGCTGCGCCCAATGGGCTGCTGCTCATCC 242  
QY 61 SerAlaSerAseuLeuTrpLeuTyThrSerArgCysArgAspAlaGlyThrGlyLeuThr 80  
DB 243 TCAGGCTGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTrpCysProGlySerGlyAlaHis 100  
DB 303 GGGCCACTGGTACCCCAACGATGGCTGAGGGTGTGTGTGCTCGAATCCGAGGCCAT 362  
QY 101 IleProLeuProProAlaProGlyGlyCysProTrpSerCysArgLeuLeuGlyValLeuGly 120  
DB 363 ATTCCCTTACCAACGCTCTGAGAGGCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTG 422  
QY 121 GlyHisLeuSerProGlyGlyValLeuLeuThrLeuProGlyGlyAlaProCysLeuLeuVal 140  
DB 423 GGGCCACTTTCCCAACGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482  
QY 141 ProArgLeuAlaGlyGlnSerCysLeuLeuAlaGlnAlaProGlyValLeuArgValGlyVal 160  
DB 483 CCACGGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542  
QY 161 ArgSerProGlyGlySerLeuGlyValValArgArgValArgAspAlaThrAlaProGly 180  
DB 543 AGGTACCAAGAGAGTCCCTGGGTGGGCTGGAAGAGAGATGTAAATACAGCCCCCAG 602

QY 181 PheGlnProProSerTyGlnAlaThrValProGlyLeuGlnGlnProAlaGlyThrProVal 200  
DB 603 TTCCAGGCCCCCACTACCAAGGCCACAGTGGCGGAGAACCAAGCAAGGACACCCCTCTT 662  
QY 201 AlaSerLeuAlaGlyAlaIleAspProAspGlyGlyValGlyValGlyValGlyValThrMet 220  
DB 663 GCATCCCTGAGGGGCACTACCAAGGCCGAGGAGTGAAGGAGGCTGCACTGAGTACCACTAG 722  
QY 221 AsnAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyVal 240  
DB 723 GATGCCCTCTTGAATAGCCGCTCCACCAAGTCTTCTTCTGAGCCCAAGTCACTGCTGCA 782  
QY 241 ValThrThrAlaGlnGlyLeuLeuAspArgGlyLeuThrTySerThrHisValPheArgValThr 260  
DB 783 GTAACCAAGCGGAGGAGGAGCTGATCTGTAGACCAAGAGCAACCACTTTCAGGGTCAAG 842  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 843 GCGCAGAGCAACGCGCAATGCCCGCAAGAGTCCCTGCTACACTCACTCTGTGTTACT 902  
QY 281 AspThrAspAspHisAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
DB 903 GACACCAATGACCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962  
QY 301 AsnLeuGlnValGlyTyGlyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 963 AACCTGAGAGGTGGCTATGAGTGTCTCATCTGAGGCCACGATGTGATGATGCCCTCTCC 1022  
QY 321 AsnAlaAsnIleLeuTyThrArgLeuLeuGlnGlySerGlySerProSerGlyValPhe 340  
DB 1023 AATCCCAATATTCGTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082  
QY 341 GluIleAspProAspArgSerGlyValIleArgThrArgGlyProValAspArgGlnVal 360  
DB 1083 GAGATCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142  
QY 361 GluSerTyGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
DB 1143 GAATCCTACCAAGCTGAGAGTGAAGCAAGTGAAGCAAGGAGTGGGACCGGGGCTCGAGGT 1202  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAspAlaProGlnPheSer 400  
DB 1203 ACCACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1262  
QY 401 GluIleAspArgTyArgValAlaGlnValArgGluAspValThrProGlyValAlaProVal 420  
DB 1263 GAGAGGCTATGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322  
QY 421 ValThrAlaSerAspAspAspAspAspGlySerAsnAlaValAlaHisTySerIleMetSer 440  
DB 1323 GTCAACGCTCGATGAGACCAAGGAGGAGCAATGCCGTGTGCTCATATGATCATGATGAT 1382  
QY 441 GlyAsnAlaArgGlyGlnPheTyThrLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460  
DB 1383 GGCATATCTGGGGACGTTTATCTGATGCTGAGCTGGAAGCTGTGATGTGTGATGCTG 1442  
QY 461 ProLeuAspTyGlnThrThrThrValGlyValThrLeuArgValArgAlaGlnAspGlyGly 480  
DB 1443 CTTTTCATATGAGACCAAGCAAGAGTACCTTACGGGTGCGAGCAAGAGATGTGTGCTG 1502  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500  
DB 1503 GGTCCCACTCTCTATATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1562  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnGlnGlnGlnGln 520  
DB 1563 AATGCCCACTCTTGTGACACACCCCTTTCAGGCTACTGTCTGAGAGAGCTCCCTTA 1622  
QY 521 GlyTyThrValIleuHisValGlnAlaIleAspAlaAspAlaGlyValAspAsnAlaArgLeu 540  
DB 1623 GGTACTCTGTTCTTCATGCTGACAGCTATGACAGCTATGCTGCTGCTGCTGCTGCTGCTG 1682  
QY 541 GluTyThrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560

```

Db 1683 GAATACGCGCTTGCGGGTGGGACATGATCCCTTTCACATCAACATGACACAGGC 1742
Qy 561 TPTILSERVALAALAGLULASPARGGLUGLUALASPHERYTSRPHGILYVAL 560
Db 1743 TGGATCTCTGTGGCTGTGGAATCGGACCGGAGGAGTTGATTTCTACAGCTTTGGGGTGA 1802
Qy 581 GLUALARGASPHISGLYTHRPROALALETUThRALASERLASEVALSERVALTHTRVAL 600
Db 1803 GAAGCTGAGACCATGGCACTTCAGACACTGACCTCGGCGAGGTGCAGCGTGAAGCTGTC 1862
Qy 601 LEUASPVALAASAPASAPASPROTHRPHETHRGLINPROGLINUTYRTHTRVALARGLEUASN 620
Db 1863 CTGGATGTCAACGACCAACATTCACCTTTACCAACAGAGTACACAGTGGGGCTCAAT 1922
Qy 621 GLUALSPALALAVAGLYTHRSERVALVALTHVALSERLAVALLASPHARGAPALAHIS 640
Db 1923 GAGGATGACACTGTGGGACACAGGTGTGACGGTGTGACGTGTGACCTGTGACCTGTGAT 1982
Qy 641 SERVALIETHRYRGLINILETHRSESGLYASNTHRARASAPARGPHESERIIETHRSER 660
Db 1983 AGGTGATCACCTACACGATCACACGAGGACATCTGGAACCGGCTTCTTCATCACACAC 2042
Qy 661 GLINSEGLYGLYGLYLEUVALSERLEUALALEUPROLEUASPTYRISLEUGLUAATGGLN 680
Db 2043 CAAGATGCTGTGGCTGTGATATCCCTTGGCCCTGACCTGACCTGACCAACTGAGCGGAG 2102
Qy 681 TYRVALLEUALAVALTHRALASERASPGLYTHRARGLINASPHRVALAGLILEVALVAL 700
Db 2103 TATGTGTGGCTGTATACCGCTCCGATGAGCACTCGGAGGACACGAGACAGATTTGGTGT 2162
Qy 701 ASNVALIETHRASVALASANTHRHISARGPROVALPHENGLINSESRHISRYRTHTRVALASN 720
Db 2163 AATGTACACGACCGCAACACCCATTCGCTCTTTCAGAGCTTCCACTATACAGTGAAT 2222
Qy 721 VALASNGLUALSPARGPROALAGLYTHRTHVALILEULLESERLATHRAASPGLUASP 740
Db 2223 GYTHAAGAGGACCGGCGGAGGACCAACGAGTGTGTGATCAACCGGACGAGTAGAGGAC 2282
Qy 741 THRGVLGLUALASVALARGILETHRTRYRPHMETGLUALSPSERILEPROGLINPHEAGTILE 760
Db 2283 ACAGGTGAGAAATGCCCATCATCCTACTTCAATGAGAGACAGATCCCGCATTCGGATC 2342
Qy 761 ASPALASPTHRGLYVALVALTHTRGLINALGLULEUASPTYRGLUALSPGLINVALSER 780
Db 2343 GATGACAGACACGGGGGCTGTCAACACCGAGCTGAGTGAATACAGAAACCAAGTGTCT 2402
Qy 781 TYRTHRLEUALAILETHRVALARGASPARAGLYLLEPROGLINYSERASPTHRTRYR 800
Db 2403 TACACCTGTGCATTTACTGTCTGGGACAAATGCAATCCCAAGAAAGTCCGACACACTAC 2462
Qy 801 LEUGLULILEUVALAASAPVALAASAPASVALPROGLINPHELEUALGASAPSETRYR 820
Db 2463 CTGGAGATCTCTGTGAAACGAGTGAATGACAAATGCCCTCAAGTCTCTGACAGACTCTAC 2522
Qy 821 GLINGLYSERVALTYRGLUALSPVALPROPHETHRSERVALLEUGLINILESERLATHR 840
Db 2523 CAAGGACGTGTCTATGAGATGTGCCACCTTCACTAGCGTCCGAGATCTCAGGACACT 2582
Qy 841 ASPHARGASPSERGLYLEUASNGLYARGVALPHETRYRTHRHEGLINGLYGLYASPARAGLY 860
Db 2583 GATCGTATTTCTGACCTTAATGAGGAGGCTCTTCAACCTTCAACAGGAGGCGACGATGGA 2642
Qy 861 ASPPGLYASPPHEILEVALGLUSERTHRSERGLYILEVALARGTHRLEUALGARGLEUASP 880
Db 2643 GAGCGTACCTTTATTTGTTAGTCAAGCTCAGGACATCGTGGAAACGCTACGAGAGGCTGAT 2702
Qy 881 ARGGLUASNVALLAGLINTYRVALLEUARGALATRYVALVALASPLYGGLYMETPROPRO 900
Db 2703 CGAGAGAACGTGGCCGAGTATGTCTTGGGGCATATGCACTGAGCAAGGAGGATGCCCA 2762
Qy 901 ALAARGTHRPROMETGLUVALIETHRVALIETHRVALLEUASPVALLASAPASAPPROVAL 920

```

```

Db 2763 GCCCGACACCTATGAGAAATGACAGTCACTGTGTTGATGTGATATGACAAATCCCTGTTC 2822
Qy 921 PHEGLUGLNASPGIUPHEASPVALLPHEVALGLUGLUALASERPROILEGLYLEUALVAL 940
Db 2823 TTTGACAGAGATGATGTTGATGTTGTTGGTGAAGAGAACAGCCCATTTGGGCTAGCCGTG 2882
Qy 941 ALAARGVALIETHRALASPPROASPPGLUGLYTHRANAALAGLILEMETYRGLINILE 960
Db 2883 GCCCGGTGTACAGCCATGACCCCATGAAAGGACCAATGCCAAGTTATGTACAGATT 2942
Qy 961 VALGLUGLYASNILEPROGLINUALPHEGLINLEUASPIIPEHSESGLYGLULEUTHRALA 980
Db 2943 GTGAGAGGCAACATCCCTGAGGCTTTCAGCTGAGCAATCTTCCGGGAGCTGACAGCC 3002
Qy 981 LEUVALASPLEUASPTYRGLUALSPARGPROGLINUTYRVALLEUVALILEGLNALATHRSER 1000
Db 3003 CTGGTGAACCTTGAACCTTACAGGACCGGCTGAGTACGTCTGTGATCAACGAGCCACGTA 3062
Qy 1001 ALAPROLEUVALSERARGLATHRVALHISVALARGLEUASPARGASAPASAPRO 1020
Db 3063 GCTCTCTGTGTAGCGCGGCTACAGTCCAGTCCGCTCTTGACCGCAATGACCAACCA 3122
Qy 1021 PROVALLEUGLYASNPHENGLINILEUPHEASNAVANTYRVALTHRASARGSESRSER 1040
Db 3123 CCAAGTGTGGGCAACTTGTGAGATCCTTTCAACAACATATGTCAACCAATGCTCAAGCAGC 3182
Qy 1041 PHEPROGLINGLYVALIILEGLYARGVALPROVALHISAPASPROVALIISERASPSERLEU 1060
Db 3183 TTCCCTGGGGGTGCAATTTGGCGAGTACCTGTCCATGACCTGATATCTCAATATGTCTG 3242
Qy 1061 THTRYSERPHEGLUALARGLYASNGLULEUASERLEUVALLEUASVALASERTHRGLY 1080
Db 3243 ACTTACAGCTTTGAGCGGGGAAATGACTCAAGCTGTGTCTCTCATATCCCTCCACGGGT 3302
Qy 1081 GLULEULYBLEUSERARGLALALEUASPARASAPARGPROLEUGLUALIILEMETSERVAL 1100
Db 3303 GAGCTGAAGCTTAAGCGCGGCACTGAGCAACACCGGCTCTGAGGCGCATATGAGCGTG 3362
Qy 1101 LEUVALSERASPGIIVALHISERVALTHRALAGLINALALEUARGVALIETHRILEILE 1120
Db 3363 CTGTGTCTGACAGCGGCTACAGGCTGACCGCCGAGTGTGCGCTGTGTATACATCATC 3422
Qy 1121 THRASPGIUMETLEUTHRTHISERILETHRLEUARGLEUGLUALSPMESERPROGLINARG 1140
Db 3423 ACCGATGAGATGCTCACCAACGACATCAAGCTGCGCTGTGAGAGCATGTCAACCGAGCGC 3482
Qy 1141 PHELEUSERPROLEUENGLYLEUPHEILEGINALAVALLAATHRLEUALATHRPRO 1160
Db 3483 TTCTGTGTCAACCTGTAGGCTCTTTCATTCAGGGGTGTGGCGGCAACGCTGTGCCACGCA 3542
Qy 1161 PROASPHISVALIIVALPHEASNVALLAGNARGASPTHRASPALPROGLYGLYHISILE 1180
Db 3543 CCGGACACAGTGTGTGTCTTCAACGTAACGAGGACACCGAGCGCCCGGGGGGCCACATC 3602
Qy 1181 LEUASNVALLSERLEUSERVALGLYGLINPROPROGLIYPROGLYGLYPROPHOLEU 1200
Db 3603 CTCAACGTAGGCTGTGTGTGGGCGAGCGCGGAGGCGCGGGGGCGCGCTTCTCTCG 3662
Qy 1201 PROSERGLUALASPLEUGLULARGLEUTHRYRLEUASNAVARGSERLEUETHRALIILESER 1220
Db 3663 CCTCTGAGAGACCTGAGAGGCGCTTATACCTCAACCGGACGCTGTGACGCGCATCTCG 3722
Qy 1221 ALAGLARGVALLEUPROPHASPARASNVILLECYBLEUARGGLUPROCYSGGLUASNVYR 1240
Db 3723 GCACAGCGGTGTGCTGCTTGAACGACCAACATCTGCTGTGGGAGGCGCTGAGAACTAC 3782
Qy 1241 METARGQSEVALSERVALLEUARGPHEASPSERSERLAPROPHETLEALASERSESR 1260
Db 3783 ATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3842
Qy 1261 VALLEUASPHARGPROILEHISAPROVALIILEGLYLYLEUARGCYAARGYAPROPROGLIYPHE 1280
Db 3843 GTGCTCTTCCGGCCCATTCACCCCGTGGAGGGGTGTGGCTGTGCTGCTGCTGCTGCTGCTT 3902

```

QY	1201	ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrIserArgProCysGlyProHis	1300
Db	3903	ACGGGTGACTACTCCAGACCCGAGGTGACCTCTGACTACCGGGCCCTGTGGCCCCCAG	3962
QY	1301	GLYArgCysAspSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyrThrGly	1320
Db	3963	GGGGCTGCGGACCCCGAGGGGGGCTACACCTGCTCTGTGTATGCTACACGGGT	4022
QY	1321	GluHisCysGluValIserAlaArgSerGlyArgCysThrProGlyValCysIysAsnGly	1340
Db	4023	GAGCAGCTGTAGGTGAGTGGCTCGCTCAGGCCGTTTGACCCCGGGTCTCTGCAAGATGGG	4082
QY	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysAspSerGlyAspPhe	1360
Db	4083	GGCACCTGTGTCAACCTGCTGTGGGGGTTTCAAGTGCATGGCCCATCTGGAGCTTC	4142
QY	1361	GluIysProTyrCysGluValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380
Db	4143	GAGAAAGCCCTTACTCCAGGTGACCAACGGAGCTCCCGCCACCTCTTCATCACTCTT	4202
QY	1381	ArgGlyLeuArgGlnArgPheHisSerPheThrLeuAlaLeuSerPheAlaThrIysGluArg	1400
Db	4203	CGCGGCTCGGCGCAGCGCTTTCACCTTCAACCTGGCCCTCTCGTTTCCCAAGAGAGGCG	4262
QY	1401	AspGlyLeuLeuLeuTyrTrasnGlyArgPheAsnGlyIysHisAspPheValAlaGlu	1420
Db	4263	GACGGGTTCGTTGTTCAAATGGCGCTTTCATAGAGAGCATGACTTCTGTGGCCCTCGAG	4322
QY	1421	ValIleGlnGluGlnValAlaGlnLeuThrPheSerAlaGlyIserThrThrValSer	1440
Db	4323	GTGATTCAGAGAGAGCTCCAGCTCACTTCTCTGCAAGGGAGTCAACCAACCGGGTCC	4382
QY	1441	ProPheValProGlyGlyValIserAspGlyGlnTrpHisThrValGlnLeuIysTyr	1460
Db	4383	CCATTTCGTGCGCGGAGGAGTCAAGTAGGCGCAGTGGCATACGGTGCATGAATATCTAC	4442
QY	1461	AsnIysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnIysValAla	1480
Db	4443	AATAAAGCCACTGTGGGTCAAGACAGGGCTCCACAGGGCCCATCAAGACAGAAAGTGGCT	4502
QY	1481	ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly	1500
Db	4503	GTGGTGAACCGTGAATGGCTGTGACACAGAGTGGCTTCCGCTTCGAGTCTGTCTGGGC	4562
QY	1501	AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyIysSerIysIysSerLeuAspLeuThr	1520
Db	4563	AACTAATCTCGTGTGCTGCCAGGGGCAACCAAGGTGGCAGCAAGAAGTCTGTGATCTGACG	4622
QY	1521	GlyProLeuLeuLeuGlyValProAspLeuProGlyIserPheProValArgMetArg	1540
Db	4623	GGGGCCCTGCTACTAGGGCGGGGTGCTGACTGCCCCGAGAGCTTCCCATCTCGAATGGGG	4682
QY	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisAlaAspMetAlaAsp	1560
Db	4683	CAGTTTCGTGGGCTGCAAGCGGAACCTGCAAGTGGACAGCCGGCACAATGACATGGCTGAC	4742
QY	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIysIysAsnValCysAspSer	1580
Db	4743	TTCAATTGCAACATGACACCGTGGCTGGCTGCCCTGCCAAGAAAGCACTGTGTGACAGC	4802
QY	1581	AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAspAlaPheSerCysGlyCys	1600
Db	4803	AACACTTGGCCAAATGGGGGCACTGTGTGACCAAGTGGAGCGCGTTCAAGCTGCAGAGTC	4862
QY	1601	ProLeuGlyPheGlyGlyIysSerCysAlaGlnGlyMetAlaAsnProGlnHisPheLeu	1620
Db	4863	CCCTCGGCTTTGGGGGCAAGAGCTGGCCCAAGAAATGGCCAAATCCACAGCACTTCTCG	4922
QY	1621	GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrLeu	1640
Db	4923	GGCAGAGCTGTGTGGCTGAGCTGTCTCTGCTGCCATCTCCCAACCCCTGTGATCTTC	4982
QY	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660
Db	4983	AGCCTCATGTCCGACCGCGCAGGCCGACCGGTCTCTGTGACAGGCCCATCCACAGGGGG	5042
QY	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGluGlyThr	1680
Db	5043	CGCAGACCAATCAACCTTACAGCTTACAGAGGGCCACGTGAATGTGAGGTGAGGGCA	5102
QY	1681	GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrpHis	1700
Db	5103	GGGCTTCAAGCCCTCTCTCTCTGCTGTGAGGCCAGGCCCGGCCAATGACGGTGACTGGAC	5162
QY	1701	HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp	1720
Db	5163	CATGCAACACTGGCACTGGAGCCAGCGGGGGGCTTGACATGCAATCTGTCTCTGCAT	5222
QY	1721	TyrGlyGlnGlnAlaArgAlaGlnIysAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer	1740
Db	5223	TATGGCAGCAGAGAGCAGAGGGGCAACTGGGGCCCCCGGCTGCATATGTGTGACCTTACG	5282
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly	1760
Db	5283	AACATTAACGTGGGGGGAATACCTGGGCAACCGCGGTGTGGCCCGTGGCTTCCGGGGC	5342
QY	1761	CysLeuGlnGlyValArgValIserAspThrProGluGlyValaAsnSerLeuAspProSer	1780
Db	5343	TGTTTTCAGGGTGTGCGGGGTGAGCATACGCACAGGGGGGTTTAAACACTGGATCCACAG	5402
QY	1781	HisGlyIserIleAsnValGluGlnIysCysSerLeuProAspProCysAspSerAsn	1800
Db	5403	CATGGGAGAGCATCAACGTGAGCAAGGCTGTAGCCCTTGACCTTGTGATCTTAAC	5462
QY	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp	1820
Db	5463	CCGTGCTCTTACACGATTAATGGACAGCAACAGACTGGGACACTATTCCTGACGCTGAT	5522
QY	1821	ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisGln	1840
Db	5523	CCAGTTACTATGTAGCAACTGTATGTGTGTGACTGAAACCCGTGTGACACAG	5582
QY	1841	SerValCysThrArgIysProSerAlaProHisGlyTyrThrCysGluCysProProAsn	1860
Db	5583	TCTGTGTGATCCCGAACCAGTGGCCCCCAATGGCTTACTCTGAGATGCTCCCCCAAT	5642
QY	1861	TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrTrpGly	1880
Db	5643	TACCTTGGGCACTGTGAGACCAAGATTGACACAGCTTGTCCCGTGGCTGTGGGA	5702
QY	1881	HisProThrCysGlyProCysAsnCysAspValIserIysGlyPheAspProAspCysAsn	1900
Db	5703	CATCCCACTGGTGGCCCAATGCAACTGTATGTACGAAAGGCTTGTGACCAAGATGGAC	5762
QY	1901	LysThrSerGlyGlyCysHisCysLeysGlnAsnHisGlyArgProProGlyIysProThr	1920



```
Db 6063 AAAGGCTCTTTGGGAGCTGTGGCCAGCTGTATGAGCAGGAGGCTGCTCCCCCA 6122
Qy 2021 AsnLeuPheAenCySerThrSer11eThrPheSerGluLeu1yGlyPheAl1GluArgLeu 2040
Db 6123 AACCTCTTCACTGACCTGACCTTCACTTCAAGAAAGGCTGCTGAGAGGCTCA 6182
Qy 2041 GluArgAenGluSerGlyLeuAapSerGlyArgSerGln1nLeuAlaLeuLeuArg 2060
Db 6183 CAGCGGAATGAGTCAAGGCTTACACTCAAGGCGCTCCAGCAGCTAGCCCTGCTCTGCGC 6242
Qy 2061 AsnAlaThrGln1n1eThrAlaGly1yThrPheGlySerAapVal1yValAla1yArg1nLeu 2080
Db 6243 AACGCCACGACAGCAGCTGCTACTTCGGCAGCAGCTCAAGTGGCCCTACCAAGCTG 6302
Qy 2081 AlaThrArgLeuLeuAla1n1eGluSerThrGln1nArg1yPheGlyLeuSerAlaThrGln 2100
Db 6303 GCCAGCGGCTGCTGGCCCAAGAGAGCAGCCAGCGGCTTGGGCTGTCTGCGCACAG 6362
Qy 2101 AapVal1n1ePheThrGlyAapLeuLeuArgVal1GlySerAlaLeuLeuAapThrAlaAen 2120
Db 6363 GACGTGACCTTCACTGAGAACTGCTGCGGGTGGGCAAGCGCTCTCTGAGACAGCCAAC 6422
Qy 2121 LysArgH1eTTPGluLeu11eGln1nThrGln1yGly1yThrAla1yTTPLeuGln1n1s 2140
Db 6423 AAGCGGCACTGGGAGCTGATCCAGCAGACAGAGGGTGGCACCGCTGCTGCTCCAGCAC 6482
Qy 2141 TyrGlnAla1yTTPAlaSerAlaLeuAlaGlnAapMetAlaGln1eThrTTPLeuSerProPhe 2160
Db 6483 TATGAGGCTTACGACAGTGGCCCTGCGCCAGAACATGCGGACACCTTACCTAAGCCCTTC 6542
Qy 2161 Thr11eVal1nThrProAap11eVal11eSerVal1ArgLeuAap1yGlyAapPheAla 2180
Db 6543 AACATGCTCAAGCCCAACTTCTCATCTCCGTAAGTCCCTTGAGCAAGGAACTTTCCT 6602
Qy 2181 GlyAla1yLeuProArgTTPGlnAlaLeuArg1yGln1nProProAapLeuGln1nThr 2200
Db 6603 GGGGCGCAAGCTGGCCCGCTACAGAGCCCTGCGTGGGAGAGCCCGCCGACCTTGAGACA 6662
Qy 2201 ThrVal11eLeuProGluSerVal1PheArgGlu1nThrProVal1ArgProAla1y 2220
Db 6663 ACAGTCACTTCTGCTGAGTCTGTCTTCAAGAGAGCCCGCCGCTGATCAGGCGCCAGGCTC 6722
Qy 2221 ProGlyGlnAlaGln1nProGluGluLeuAlaArg1nArgGln1nArgH1eProGluLeu 2240
Db 6723 CCCGAGAGGCGCCAGAGGCAAGAGCTGGCAGCGGCAAGCGCACCCGGAAGCTG 6782
Qy 2241 SerGln1yGlnAlaValAlaSerVal11e1yTTPArgThrLeuAlaGlyLeuLeuPro 2260
Db 6783 AGCCAGGCTGAGGCTGGCCAGCTCATCTACCCGACCCCTGGCGCTACTGCTC 6842
Qy 2261 HisAen1yTTPAapProAap1yAapSer1eLeuArgVal1Pro1yArgPro11e1eAenThr 2280
Db 6843 CATTACTTACCTTACCAAGCAGCTTGAAGTCCCAAGCGCCCAATCATCAACACA 6902
Qy 2281 ProVal1Ser11eSerVal1n1eAapAapGln1nLeuLeuProArgAlaLeuAap1yS 2300
Db 6903 CCCGAGTGAAGCATCAGGCTCATGATATGAGAGCTTTCGCCCGGCGCTTGACAAA 6962
Qy 2301 ProVal1ThrVal1GlnPheArgLeuLeuGlu1nThrGln1nArgThr1yLeuPro11eCyVal1 2320
Db 6963 CCCGTCACGCTGACGCTTCCGCTGAGACAGAGGAGCGGACCAACCCCATCTGCTGTC 7022
Qy 2321 PheTTPAenH1eSer11eLeuVal1SerGlyThrGln1yTTPSerAlaArgGlyCyGln1n 2340
Db 7023 TTCTTGAAACATTCATCTGCTCACTGAGCAGAGTGGCTGCTGCGCCAGAGGCTGTAAA 7082
Qy 2341 Val1Val1PheArgAapGluSerH1eVal1SerCyGln1nCyAapH1eMetThrSerPheAla 2360
Db 7083 GTGCTCTTCCGCAATGAGAGCAGCTCAGCTCCAGTGCACCATGACGAGCTTCCCT 7142
Qy 2361 Val1LeuMetAapVal1SerTTPArgGluAenGlyGlu11eLeuProLeu1yThrLeuThr 2380
Db 7143 GTGCTATGAGACGTTTCTCGCGGAGAAATGAGGAGATCTCGCACATGAAGACATGACA 7202
Qy 2381 TyrValAlaLeuGlyVal1ThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeu 2400
Db 7203 TACGTGCTCTAAGTGTGACCTTGGCTGGCTGCTGTCTGCTCACTTCTTCTTCACTCTC 7262
Qy 2401 LeuArg11eLeuArgSerAenGln1n1eGly11eArgArgAenLeuThrAlaAlaLeuGly 2420
Db 7263 TTGGGTACTTGGCGCTTCAACCAACAGGCAATCCAGCTAATCTGACAGCTGCCCTGGCC 7322
Qy 2421 LeuAlaGlnLeuVal1PheLeuLeuGly11eAapGlnAlaAapLeuProPheAla1ySerThr 2440
Db 7323 CTGGCTCAGCTGGCTTCTCTTGGGAATCAACAGGCTGACCTCCCTTTCCTGACACA 7382
Qy 2441 Val11eAla11eLeuLeuH1ePheLeuTTPLeu1yCysThrPheSerTTPAlaLeuGln1n 2460
Db 7383 GTCATTCACATCTCGTCACTTCTGTAATCTTGCACCTTTCCTGGGCTGTGAGAG 7442
Qy 2461 AlaLeuH1eLeuTTPArgAlaLeuThrGluVal1ArgAapVal1AenThrGlyProMetArg 2480
Db 7443 GCCTTGACCTGTACCGGGGACATCACTGAGTGGCAATGTCAACACCGGCCCATGCTGC 7502
Qy 2481 PheTTPThrMetLeuGlyTTPGlyVal1ProAlaPhe11eThrGlyLeuAlaVal1yLeu 2500
Db 7503 TTCTACTACATGCTGGGCTGGGCGTGGCTGCTTCATCAAGGGCTAGCCGTGGGCTG 7562
Qy 2501 AapProGluGlyTTPArg1yAapProAapPheCyTTPLeuSer11eTTPAapThrLeu11e 2520
Db 7563 GACCCCAAGGAGCTACGAGAAACCTGACTTGTGCTGCTCCATCTAATGACAGCTATC 7622
Qy 2521 TTPSerPheAlaGlyProValAlaPheAlaVal1SerMetSerVal1PheLeuTTP11eLeu 2540
Db 7623 TGAAGTTTGTGCGCCGGGAGCTTGGCTTGGCTTCATGATGATGCTTCTCTGATCATCTG 7682
Qy 2541 AlaAlaArgAlaSerCyValAlaGlnArgGln1yPheGln1yGly1yProVal1Ser 2560
Db 7683 GCGGCGCGGCTCTCTGTGCTGCGCCAGCGGCTTTTGAAGAAAGTCCCTGTCTCG 7742
Qy 2561 GlyLeuGln1nProSerPheAlaVal1LeuLeuLeuSerAlaThrTTPLeuLeuAlaLeu 2580
Db 7743 GGGCTGAGGCTCTCTTGGCGCTCTCTGTCTGTGAGCGGCAAGTGGCTGGCACTG 7802
Qy 2581 LeuSerVal1AapSerAapThrLeuLeuPheH1eTTPLeuPheAlaThrCyAapCyAl1e 2600
Db 7803 CTCTGTCAACAGCAGCACCTCTCTTCCATCACTCTTGTACTGCAATTTGATC 7862
Qy 2601 Gln1yProPhe11ePheLeuSerTTPVal1Val1LeuSer1yGln1nVal1GlyValAlaLeu 2620
Db 7863 CAGGCGCCCTTCACTTCTCTCTATGTGTGTGCTTGAAGAGGCTCCGAAAGCACTC 7922
Qy 2621 LysLeuAlaCySerAArg1yProSerProAapProAlaLeuThrThrLysSerThrLeu 2640
Db 7923 AAGCTTGTGCAAGCCCAAGCCCAAGCTGACCTTCTGTACACCAAGTCCACCTG 7982
Qy 2641 ThiserSerTTPAapCyProSerProTTPAlaAapGlyArgLeuTTPArg1nProTTPGly 2660
Db 7983 ACTGCTGTCAACATGCCCAAGCCCTTACAGATGGCGGCTGTACAGCCCTTACGGA 8042
Qy 2661 AapSerAlaGlySerLeuH1eSerThrSerArgSerGly1yLysSerGln1nProSerTTP11e 2680
Db 8043 GACTGCGCGGCTCTTGCACAGACCAAGTCCCTGGGCAAGAGTCAAGCCCAAGTATATC 8102
Qy 2681 ProPheLeuLeuArgGln1nSerAlaLeuAapProGlyGln1n1yProProGly1yLeuGly 2700
Db 8103 CCTTCTTGTGAGGAGAGAGTCCGACATGACCTTGGCCAAAGGCGCCCTGTGGCTGGG 8162
Qy 2701 AapProGlySerLeuPheLeuGln1yGlnAapGln1nH1eAapProAapThrAapSer 2720
Db 8163 GATCAGAGAGCTGTCTCTGGAAGTCAAGACCAAGACATGATCTTGCACAGCACTCC 8222
Qy 2721 AapSerAapLeuSerLeuGln1nAapAapGln1nSerGly1ySerTTPAlaSerThrH1eSerSer 2740
Db 8223 GACAGTACCTTGTCTTGAAGAGACCAAGTGGCTCTTATGCTTACCACTCATCA 8282
```





QY 361 GIUSerTYrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProAspSer 380  
 DB 1081 GATCTTACAGCTGACGGTAGAGGACAAAGGTACGAGGTGCGGACCCGGGCTCTCGAGCT 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 DB 1141 ACCACAGCCGCTGTTTCTCTTCTGTGAGAGATGACAAATGATATATGCCCCCAAGTTTACT 1200  
 QY 401 GlnTyArgTYrValValGlnValArgGluAspValThrProGlyValAlaProValLeuArg 420  
 DB 1201 GAGAAGCGTATGTGTCCAGGTGAGGAGAGATGTACTCCAGGGGCCCACTATCTCCCA 1260  
 QY 421 ValThrAlaSerAspArgAspGlySerAspAlaValAlaHisTYrSerIleMetSer 440  
 DB 1261 GTACACAGCTCGGATGACAAAGGAGACCAATGCGGTGTGACTATAGCATCATGAGT 1320  
 QY 441 GAlaAspAlaArgGlyGlnPheTYrLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460  
 DB 1321 GGCAATGCTCGGGGACAGTTTATCTGAGATGCCACAGACTGAGCTCTGATGTGTGAGAC 1380  
 QY 461 ProLeuAspTYrGlnThrThrArgGlyTYrThrLeuArgValArgAlaGlnAspGlyGly 480  
 DB 1381 CCTCTTGACTATAGAGACGACCAAGAGTACACCTTACGGGTGCGAGCACAGATGTGAC 1440  
 QY 481 ArgProProLeuSerAspValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500  
 DB 1441 CGTCCCCCACTCTATGTCTCTGTGCTGTGACAGTACAGTCTGTGATATCAACGAC 1500  
 QY 501 AspAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 DB 1501 AATGCCCATCTTCTGTGACACCCCTTTCAGGCTACTGTCTGTGAGGCGTCCCTTA 1560  
 QY 521 GlyTYrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeu 540  
 DB 1561 GGCTACTGCTTCTCATGTCCAGGCTATCGAGCTGATGCTGTGTGACAAATCCCGCTCG 1620  
 QY 541 GlnTYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAspAsnGlyTYrGly 560  
 DB 1621 GAATACCGCTGTGTGAGGAGCAATGACTTCCCTTCCATCAACATATGCGACAGGC 1680  
 QY 561 TPPIleSerValAlaAlaGlnLeuAspArgGlnGluValAspPheTYrSerPheGlyVal 580  
 DB 1681 TGGATCTCTGTGGCTCTAACTGACCGGAGAGATGATTTCTACAGCTTGGGGTA 1740  
 QY 581 GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 DB 1741 GAAGCTCGAGACCAATGGCACTCCAGCACTCACTGGCTGGCCAGTGTGACTGTG 1800  
 QY 601 LeuAspValAspAspAspAspProThrPheThrGlnProGlnTYrThrValArgLeuAsn 620  
 DB 1801 CTGGATGTCAACGACAAACATCCAACTTTACCCCAACAGATACCAAGTGGCGCTCAT 1860  
 QY 621 GluAspAlaAlaValGlyTYrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 DB 1861 GAGATGACGCTGTGGGACCAAGCTGTGTGACGCTGTGACCCGTATGTCTCAT 1920  
 QY 641 SerValIleThrTYrGlnIleThrSerGlyAspThrArgAspAspPheSerIleThrSer 660  
 DB 1921 AGGTGATCATCTTACAGATCAACAGTGGCAATATCTGAAACCGCTTCCATTCACAGC 1980  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTYrIleLeuGlnArgGln 680  
 DB 1981 CAAAGGTGTGTGGGTGTATCTCTTGTGCTGCACTGACATCAAACTTAAAGGGGCG 2040  
 QY 681 TYrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 DB 2041 TATGTGTGTGCTGTATCCGCTCCATGTGCACTCGGCAAGACAGGCAAGATGTGGTG 2100  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYrThrValAsn 720  
 DB 2101 AATGTACCGAGCGCAACCCATGTCTGTCTTCAAGAGCTCCCATATATCAGTGAAT 2160

QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 DB 2161 GTTAATGAGACCGGCGCGGACGACCAAGGTGTGTATCAAGCGGACCAAGATGAGAC 2220  
 QY 741 ThrGlyGluAspAlaArgIleThrTYrPheMetGluAspSerIleProGlnPheArgIle 760  
 DB 2221 ACAGGTGGAATGCGCGATCATCTTCAATGAGGACACACATCCCGCATTCGCGATC 2280  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeuAspTYrGluAspGlnValSer 780  
 DB 2281 GATGAGACACAGGGGCTGTATCCAGGCTAGCTGAGCTAGCTAGACCAAGCACTGTCT 2340  
 QY 781 TYrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnTySerAspThrThrTYr 800  
 DB 2341 TACACCTGTGGCACTTACTGTGCGGACAAATGGCAATTCGCCAAGTCCGACCACTTAC 2400  
 QY 801 LeuGlnIleLeuValAspAspValAspAspAspAlaProGlnPheLeuArgAspSerTYr 820  
 DB 2401 CTGAGATCTGTGTGAAGAGCTGAATGACAAATGCCCTCAGTTCCTGCGAGACTCTTAC 2460  
 QY 821 GlnGlySerValTYrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 DB 2461 CAGGCACTGTCTATGAGATGTGTCCACCTTCACTAGCTGTGTGTGATCTTCAAGCACT 2520  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTYrThrPheGlnGlyGlyAspAspGly 860  
 DB 2521 GATCGTATCTGTGACTTAATGAGCGAGGTCTTCACTTCAAGGAGGCGAGATGGA 2580  
 QY 861 AspGlyAspPheIleValGlnSerThrSerGlyIleValAlaGlyThrLeuArgArgLeuAsp 880  
 DB 2581 GACGGTACTTATGTGTGATGCCAGTCCAGGCAATCGGCAACGCTACGAGGCTGGAT 2640  
 QY 881 ArgGluAsnValAlaGlnTYrValLeuArgAlaTYrAlaValAspGlyGlyMetProPro 900  
 DB 2641 CGAGGAAAGTGTGGCCAGTATGTCTTGGGCGCATATGAGTGTGACAGGGGATCCCCCA 2700  
 QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAspAspAspProVal 920  
 DB 2701 GCCCGCACCTTATGAGGAATGACATCATGTGTGTGATGTGAATGACAAATCCCCCTGTC 2760  
 QY 921 PheGlnGluAspGlyIlePheAspValPheValGlnGluAsnSerProIleGlyLeuAlaVal 940  
 DB 2761 TTTGAGCGAGTGTGATGT 2820  
 QY 941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTYrGlnIle 960  
 DB 2821 GCCCGGATCAAGCACTGACCCCGAGAGGACCAATGCCAGATATGTACAGATT 2880  
 QY 961 ValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePheSerGlyGlnLeuThrAla 980  
 DB 2881 GTGAGGGCAACATCCCTGAGGTCTTCCAGCTGACATCTTCTCCGGGAGACTGACAGCC 2940  
 QY 981 LeuValAspLeuAspTYrGluAspArgProGlnTYrValLeuValIleGlnAlaThrSer 1000  
 DB 2941 CTGTGATCTTGACTAGAGAGACCGGCTGAGTACGTCTGTGTATCCAGGCACTGTCA 3000  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspPro 1020  
 DB 3001 GCTCCTGTGTGAGCCGGGCTTACAGTCCAGTCCGCTCTTGAACCGCAATGACACCA 3060  
 QY 1021 ProValLeuGlyAspPheGlnIleLeuPheAspAsnTYrValThrAspArgSerSer 1040  
 DB 3061 CCAAGTGTGGGCAACTTGAATCTTTTCAACACATATGTACCAATCGTCAAGCAGC 3120  
 QY 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 DB 3121 TTCCCTGGGGGTGCAATTGGCGAGTACCTGCACTGATATCTCATATGATGTG 3180  
 QY 1061 ThrTYrSerPheGlnuArgGlyAsnGlnLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 DB 3181 ACTTACAGCTTGTGAGCGGGAATGAACTCAGCTGTGCTGTCAATGTCTCCACGGGT 3240  
 QY 1081 GluLeuTyIleuSerArgAlaLeuAspAspAsnArgProLeuGlnAlaIleMetSerVal 1100

3241 GAGCTGAAGCTAAGCGCGCATGAGCAACAACCGGCTCTGAGAGCATATGAGCGTG 3300  
1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
3301 CTGGGTGTCAGAGGGGTACAGAGCGTGACCGGCCAGTGGCGCTGGTGTGACATCATC 3360  
1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
3361 ACCGATGAGATCTCTACCCACAGCATCAGCTGCGCTGAGAGACATGTACCCAGCGC 3420  
1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
3421 TTCCTGTCAACCACTGAGGCGCTTCATCCAGCGGTGGCGGCACGCTGGCCACAGCCA 3480  
1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyValHisIle 1180  
3481 CCGAGACACGTGTGTGTCTTCAACGTACAGCGGAGACCGAGCGCCCCCGGGGGCCACATC 3540  
1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyValGlyProProPheLeu 1200  
3541 CTCAACGTGAGCTGTGGTGGTGGCCAGCGCCAGGGGCCGGGGCGCGCCCTTCTCTG 3600  
1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
3601 CCTCTGAGAGACTGTCAGAGGCGCCTATACCTCAACCGCAGCGCTGACGCGCATCTCG 3660  
1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGlnProCysGluAsnTyr 1240  
3661 GCACAGCGGTGTCTCCCTTCACACACATCTGTGGGAGACCCCTGCAGAACATCAC 3720  
1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
3721 ATGGCTGT 3780  
1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
3781 GTGCTCTTCCGGCCCATCCACCCCGTGGAGGGGCTGCGTGGCGCTGCCCGCGGCTTC 3840  
1281 ThrGluAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
3841 ACGGGTGACTACTGTGAGAGCCGAGGGTGAACCTCTCTACTCGCGGCCCTGTGGCCCCAC 3900  
1301 GlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
3901 GGGCGCTCGCGAGCGCGGAGGGCGGCTACCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 3960  
1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGly 1340  
3961 GAGCACTGTGAGT 4020  
1341 GlyThrCysValAsnLeuLeuValGlyGlyPheIleCysAspCysProSerGlyAspPhe 1360  
4021 GGCACCTGTGTACCTGT 4080  
1361 GlnIleProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
4081 GAGAGCCCTTACTGTCAAGTGTACACGCGCACTTCCCGCCCACTCTTATCACTCTTT 4140  
1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGluArg 1400  
4141 CGCGGCTCGCGCACGCTTCCACTTCACTCCGCGCTCTGTGTGTGTGTGTGTGTGTGTGT 4200  
1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIleHisAspPheValAlaLeuGln 1420  
4201 GACGGGT 4260  
1421 ValIleGlnGluGlnValAlaGlnLeuThrPheSerAlaGlyIleSerThrThrValSer 1440  
4261 GTGATCCAGAGAGAGTCCAGCTCACTTCTCTGTGAGGGAGTCAACACACGCTGTCTC 4320  
1441 ProPheValProGlyValValSerAspGlyGlnThrPheThrValGlnLeuIleTyrTyr 1460

4321 CCATTGTCGCCCGAGAGATGATGATGCCATAGCGTGCATAGCGTGAATACTACTAC 4380  
1461 AsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIleValAla 1480  
4381 AATAGCCACTGT 4440  
1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGluSerValLeuGly 1500  
4441 GTGTGTACCGT 4500  
1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlySerIleCysSerLeuAspLeuThr 1520  
4501 AACTACTCTGT 4560  
1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
4561 GGGCCCCCTGTATAGGGGGGT 4620  
1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
4621 CAGTTCTGT 4680  
1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIleCysValAsnValCysAspSer 1580  
4681 TTCAATGCCACATATGGACCGGT 4740  
1581 AsnThrCysHisAsnGlyGlyTyrCysValAsnGlnThrAspAlaAspSerCysGluCys 1600  
4741 AACCTTGCCACATATGGGGGACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
1601 ProLeuGlyPheGlyGlyTyrSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
4801 CCTCTGGCTTTGGGGGCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4860  
1621 GlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640  
4861 GGCAGAGCCTGT 4920  
1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
4921 AGCTCATGTTCGACGCGCGCAGCGCGACGAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 4980  
1661 ArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGlyGlyThr 1680  
4981 CGCAGACATCATCCTTACAGCTTACAGAGGGCCACTGATGTGTGTGTGTGTGTGTGTGTGT 5040  
1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyValArgAlaAsnAspGlyAspTyrHis 1700  
5041 GGGCTTCAAGGCTGTCTCTCTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100  
1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaAlaIleLeuSerPheAsp 1720  
5101 CATCACAAGCTGT 5160  
1721 TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleSer 1740  
5161 TATGGGAGAGAGAGAGAGAGGGAACCTGGGCCCGCGCTGTGTGTGTGTGTGTGTGTGTGT 5220  
1741 AsnIleThrValGlyIleProGlyProAlaGlyValAlaIleArgGlyPheArgGly 1760  
5221 AACATTAACAGTGGCGGAAATACCTGTGGCCAGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 5280  
1761 CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer 1780  
5281 TGT 5340  
1781 HisGlyIleSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
5341 CATGGAGAGACATCAACGT 5400  
1801 ProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyrSerCysAsp 1820  
5401 CCGTGTCTGTGTATAGCTATGT 5460

[illegible]

QY	2181	GLYAlaLeuValSerProAGYTrGValaLeuArgLysGluGlnProPheAspLeuGluThr	2200
Db	6541	GGGGCCAAAGCTGGCCCCCTCAAGAGCCCTGGGTGGAGAGCAAGCCCCCGAGCTTGAACA	6600
QY	2201	ThrValIleLeuProGluSerValPheArgGluThrProValValArgProAlaGly	2220
Db	6601	ACAAGTCAATTCGTGGAGTCTGTCTTCAAGAGACGCCCCCGTGGTCAAGGCCCGCAAGGC	6660
QY	2221	ProGlyValIleArgGlnLysProGluGlnLeuAlaArgArgGlnAlaArgAlaSerProGluLeu	2240
Db	6661	CCCGAAGAGGCCCAAGAGCCCAAGAGACCTGGCAAGGAGCAAGCAAGCAAGCCGAGAGCTG	6720
QY	2241	SerGlnGlyGluAlaValAlaSerValIleIleYrArgThrLeuAlaGlyLeuLeuPro	2260
Db	6721	AGCCAGGGGTGAGGCTGTGGCCAGGCTATCATCTACCGCAACCCCTGGCGAGGCTACTGGCT	6780
QY	2261	HisLeuTYrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr	2280
Db	6781	CATACCATATGACCTGTGCAAGCCGAGCTTGAAGTCCCAAGCCGCCCATCATCAACACA	6840
QY	2281	ProValValSerIleSerValHisAspAspGlnGluLeuSerProArgAlaLeuAspLys	2300
Db	6841	CCCGTGGTGAACATCAAGCGTTCATGATGATGAAGAGCTTCTGGCCCCGGGCCCTGGACAA	6900
QY	2301	ProValThrValGlnPheArgLeuLeuGluThrGluLysArgThrLysProIleCysVal	2320
Db	6901	CCCGTCAACGGTGCAGTTCCCGCTGTCTGAGACAGAGAGCGAGACCAAGCCCATGTGTGTC	6960
QY	2321	PheThrAsnHisSerIleLeuValSerGlyThrGlyYrTrpSerAlaArgGlyCysGlu	2340
Db	6961	TTCTGGAAACCAATTCATCTGTGTCAAGTGGCAAGGTGCTGTGTGGCCAGAGGCTGGAA	7020
QY	2341	ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla	2360
Db	7021	GTGGTCTTCCGCAATGAGAGCCATGTCAAGTGCAGTGGCCAGTGAACCAATGAGAGGCTTGCT	7080
QY	2361	ValLeuMetAspValSerArgArgLysArgGlyLysIleLeuProLeuLysThrLeuThr	2380
Db	7081	GTGTCTATGACAGCTTTCTCGGCGGGAATGGGAGATCTTGCCACATGAAGACACTACA	7140
QY	2381	TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu	2400
Db	7141	TACGTGGCTCAAGTGTCACTTGCGTGGCTGCTTGTGTCACTTCTTCTTCTCACTCTC	7200
QY	2401	LeuArgIleLeuArgSerArgGlnHisGlyYrLeaArgArgAsnLeuThrAlaAlaLeuGly	2420
Db	7201	TTGGGTATCTCGCGCTCAACCAACGCGATCCGAGTAACTGACAGGCTGCCCTGGGCG	7260
QY	2421	LeuAlaGlnLeuValPheLeuLeuGlyYrIleAsnGlnAlaArgLeuProCysPheAlaCysThr	2440
Db	7261	CTGGCTAAGCTGGCTTCTCTCTCGGAGATCAACAGGCTACCTCCCTTTTGCTTGACA	7320
QY	2441	ValIleAlaIleLeuLeuHisAspLeuTYrLeuCysThrPheSerTrpAlaLeuLeuGlu	2460
Db	7321	GTCAATGCAATCCGCTGCACTTCTCTGTACCTTGACACTTTTCTGGGCTCTGGTGAAG	7380
QY	2461	AlaLeuHisIleuTYrArgAlaAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Db	7381	GCCCTTGACCTGTACACCGGGAACCTCAAGAGTGGCGATGTCAACACCGGCCCATATGCG	7440
QY	2481	PheTYrTYrMetLeuGlyTYrGlyValProAlaPheIleThrGlyLeuAlaValAlaGlyLeu	2500
Db	7441	TTTCACTCAATGCTGGGCTGGGGCGTGGCTGTCTTATCAACAGGCTACAGCGTGGGCTGG	7500
QY	2501	AspProGlnGlyTYrGlyAsnProAspPheCysTrpLeuSerIleYrAspThrLeuIle	2520
Db	7501	GACCCCAAGGGCTACAGGGAACCTCTGACTTCTGGCTGTCTCCATCTATGACACGCTCATC	7560
QY	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTYrIleLeu	2540
Db	7561	TGAGATTTTGTGGCCCGGTGGCTTTGGCGTCTCATATGATGTCTTCTCTGTACATCTTG	7620
QY	2541	AlaAlaArgAlaSerCysAlaAlaGlnArgGlnLysPheGluLysLysGlyProValSer	2560

```

Db      7621 GGGGCCCCGGGCTCTGCTCTGCGCAGCGGCGGCTTGGAGAGAAAGCTCTGCTCG 7680
Oy      2561 G1yleuGlnProSerPheAlaValLeuLeuLeuSerAlaThrProLeuLeuAlaLeu 2580
Db      7681 GGCTTCAGACCTCTCTCGCGCTCTCTCTGCTGCTAGCGCCAGCGGCTGCTGCGACTG 7740
Oy      2581 LeuSerValAsnSerAspThrLeuLeuPheH1sTyrLeuPheAlaThrCybaAncy1le 2600
Db      7741 CTCTGCTGCAAGGAGACCTCTCTCTCACTTCACTTCTTGTCTACCTGCAATTGCAATC 7800
Oy      2601 G1nglyProPheH1lePheLeuSerTyrValValLeuSerLyGluValArgLyAlaLeu 2620
Db      7801 CAGGGGCCCCCTCATCTTCTCTCTCAATGCGCTTACAGAGAGAGGTCCGGAAGACTTC 7860
Oy      2621 LybLeuAlaCySeSerArgLyAsProSerProAspProAlaLeuThrThrySerThryLeu 2640
Db      7861 AAGCTTGCTGCAAGCGGCAAGCCCTGAGCCCTGCTCTGACCAACAGTCCACCTG 7920
Oy      2641 ThrSerSerTyrAsnCyAsProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660
Db      7921 ACCTGCTCTTACAGCTGCGCCAGCCCTTACGAGAGTGGCGGCTGTACAGCCCTTACGGA 7980
Oy      2661 AspSerAlaGlySerLeuH1sSerThrSerArgSerGlyLySeSerGlnProSerTyr1le 2680
Db      7981 GACTGGCGGCTCTCTGCAAGCAGCACTCCCTCGGCGAGAGTACAGCCAGCTTACATC 8040
Oy      2681 ProPheLeuLeuArgGlyLeuSerAlaLeuAsnProGlyGlnGlyProProGlyLyLeuGly 2700
Db      8041 CCTCTCTTCTGAGAGAGAGAGTCCGACTGAAACCTTGGCCAGAGGCGCCCTGCGGGG 8100
Oy      2701 AspProGlySerLeuPheLeuGlyGlnAspGlnGlnH1sAspProAspThrAspSer 2720
Db      8101 GATCCAGGAGAGCTGCTTCTGAGAGTCAAGACAGAGCATGATCTTGAACAGCACTCC 8160
Oy      2721 AspSerAspLeuSerLeuGlyLeuAspArgGlnSerGlySerTyrAlaSerThrH1sSerSer 2740
Db      8161 GACAGTGAACCTGCTCTTGAAGAGACAGCAAGTGTCTCTATGCTCTTACCCACATCATCA 8220
Oy      2741 AspSerGlyGlyGlyProGlyLyLeuAlaProTyrProGlyAspPheGlyThrThralyGly 2760
Db      8221 GACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280
Oy      2761 TrpAspSerLeuLeuGlyProGlyAlaGlyArgLeuProLeuH1sSerThrProLyAsp 2780
Db      8281 TGGGATAGCTGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8340
Oy      2781 GlyGlyProGlyProGlyLyLeuAlaProTyrProGlyAspPheGlyThrThralyGly 2800
Db      8341 GGGGGCCAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8400
Oy      2801 SerSerGlyAsnGlyAlaProGlyGlyArgLeuArgLeuArgLeuArgLeuArgLeu 2820
Db      8401 AGTATGCGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8460
Oy      2821 G1nglySerLeuGlyProLeuProGlySerSerAlaGlnProH1sLyS----- 2836
Db      8461 CAGGGGCTCCCTAGAGGCTCTTCTGCAAGCTTCTGCGCAAGCTCAAAAGGTGAGTGGGGC 8520
Oy      2836 ----- 2836
Db      8521 ACCCCAGCTGCGAGCTCCCTAGTACAGACCTCATACCTCATCTCTCTGCGCG 8580
Oy      2837 -----G1yleuLybLySeCyLeuProThr1le 2847
Db      8581 CACTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8640
Oy      2848 SerGlyLySerSerLeuLeuArgLeuProLeuGlnGlyGlnGlySerSerArgLy 2867
Db      8641 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8700
Oy      2868 SerSerAlaSerGlySerArgLyGlyProProProArgProProArgGlnSer 2887

```

```

Db      8701 TCCTCCGCTAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8760
Oy      2888 LeuGlnGlyGlnLeuAsnGlyValMetProLeuAlaMetSer1leLyAlaGlyThrVal 2907
Db      8761 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8820
Oy      2908 AspGluAspSerSerGlySerGlyLeuPheLeuPhePheAsnLeuH1s 2923
Db      8821 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8888

RESULT 5
LOCUS      AX367102          9321 bp    DNA          linear    PAT 16-FEB-2002
DEFINITION Sequence 21 from Patent WO0198354.
ACCESSION  AX367102
VERSION     AX367102.1 GI:18855304
KEYWORDS
SOURCE
  ORGANISM  Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 Griffin,J.A., Kallick,D.A., Tribouley,C.M., Yue,H., Nguyen,D.B.,
    Tang,Y.T., Lal,P., Policky,J.L., Azimzal,Y., Lu,D.A., Graul,R.,
    Yao,M.G., Burford,N., Hafalia,A.J., Baughn,M.R., Bandman,O.,
    Patterson,C., Yang,J., Xu,Y., Warren,B.A., Ding,L. and
    Sanjwal,M.S.
  Receptor
  Patent: WO 0198354-A 21 27-DEC-2001;
  Incyte Genomics, Inc. (US)
FEATURES
  source
    1..9321
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /note="Incyte ID No: 6977010CB1"
BASE COUNT  1780 a 3008 c 2776 g 1757 t
ORIGIN
Alignment Scores:
Pred. No.:      0          Length:      9321
Score:          15279.00    Matches:      2893
Percent Similarity: 98.44%   Conservative: 2
Best Local Similarity: 98.37% Mismatches: 14
Query Match:    98.29%      Indels:      32
DB:              Gaps:      8

US-09-916-849a-3 (1-2923) x AX367102 (1-9321)
Oy      1 MetArgSerProAlaThrGlyValProLeuProThr---ProProProProLeuLeu 19
Db      223 ATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 282
Oy      20 LeuLeuLeuLeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCyArgSer 39
Db      283 CTGTGTGCTGCTGCTGCTGCGCGCCACTATGAGAGACCAAGTGGAGGCTGTGCTTCC 342
Oy      40 LeuGlySerArgGlyArgGlySerSerGlyAlaGlyAlaProMetGlyTyrPleuCyPro 59
Db      343 TTGGGGTCCAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
Oy      60 SerSerAlaSerLeuLeuLeuTyrPleuTyrThrSerArgCybaArgAspAlaGlyThrGlyLeu 79
Db      403 TCCCTCAAGCTGAGACCTCTGCTCTTACAGAGCGCTGAGAGATGCGGAGACTGAGACTG 462
Oy      80 ThrGlyH1sLeuValProH1sH1sAspGlyLeuArgValTyrCyAsProGlySerGlyAla 99
Db      463 ACTGGCACTGTGATCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
Oy      100 H1s1leProLeuProProAlaProGlyGlyCyProThrSerCybaArgLeuLeuGly1le 119
Db      523 CATATTCCTTACCAACAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582

```

QY 120 GYGLVHIEUSErProGInglYlYsLeuThrLeuProGluGluHIErProCYsLeuYs 139  
DB 583 GGAGGCCACTTTCCTCCACAGGGCAAGCTCACCTCCGAGGAGCACCGGTGCTTAAG 642  
QY 140 AIAProArgLeuArgCYsGInSerCYsLYsLeuAlaGlnAlaProGluYLeuArgAlaGly 159  
DB 643 GCTCCAGGCTCAGAGTCCAGCTCTCCAGACTGGCACAGGCCCGGGGTCCAGGGCAGGG 702  
QY 160 GIUArgSerProGluGluSerLeuGlyGlyArgArgLYsArgAsnValAsnThrAlaPro 179  
DB 703 GAAAGGTCACCAAGAAAGTCCCTGGGTGGCGCTCGAAAAGAAATTAAATACAGCCCC 762  
QY 180 GlnPheGlnProProSerTYrGlnAlaThrValProGluAsnGlnProAlaGlyThrPro 199  
DB 763 CAGTTCACAGCCCCCAGCTACCAAGGCCACAGTCCGGAGAACCAAGCCACAGGCACTT 822  
QY 200 ValAlaSerLeuArgAlaIleAspProAspGluGluAlaGlyArgLeuGluTYrThr 219  
DB 823 GTTGCAATCCCTGAGGGGCATCCAGCCGAGAGGTGAGGCAAGTCACTGAGTGAACACC 882  
QY 220 MetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGly 239  
DB 883 ATGGAGTCCCTCTTATAGCCGCTCCAAACAGTCTTCTCCCTGAGCCCAAGTCACTGCT 942  
QY 240 AlaValThrThrAlaGluGluLeuAspArgGluThrLYsSerThrHisValPheArgVal 259  
DB 943 GGAGTACCAACAGCCCAAGAGGTGAGTCTGTAGAACCAAGACCAACCTCTTCAAGGCTC 1002  
QY 260 ThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuVal 279  
DB 1003 AGGGCCAGAGACCAAGGACAGCCCAAGAGTCCCTGCTACCTCACTCAACTTGGT 1062  
QY 280 ThrAspThrAsnAspHisAspProValPheGluGlnGluTYrLYsGluSerLeuArg 299  
DB 1063 ACTGACACCAATGACATGACCTGTGTTGAGACAGAGATGACAGAGAGAGCTCAAG 1122  
QY 300 GluAsnLeuGluValGlyTYrGluValLeuThrValArgAlaThrAspGluAspAlaPro 319  
DB 1123 GAGAACCTGAGAGGTGCTATGAGGTGCTCACTGTCAAGGGCCAGAGTGTGATGCCCC 1182  
QY 320 ProAsnAlaAsnIleLeuTYrArgLeuLeuGluGlySerGlyLYsSerProSerGluVal 339  
DB 1183 CCCAATGCCAATTTCTGACCGCTGCTGAGAGGATCTGGGGGACAGCCCTCTGAAGTC 1242  
QY 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
DB 1243 TTGGAGATGAGCCTCGCTCTGGGGTGAATCCGAACCGTGGCCCTGTGATGGGAGAG 1302  
QY 360 ValGluSerTYrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379  
DB 1303 GTGGAAATCTTACCAAGCTGAGGTGAGGACAGTGAACAGGCTCGGGACCCGGGTCTCCG 1362  
QY 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPhe 399  
DB 1363 AGTACACAGCCGCTGTTTCTTCTGAGAGAGTGAATGATTAATGCCCCCAAGTTT 1422  
QY 400 SerGluLYsArgTYrValValGlnValArgGluAspValThrProGlyValaProValLeu 419  
DB 1423 AGTGAAGAGGCTATGTGTCCAGGTGAGGAGAGATGTGATCTCAAGGGGCCCAAGTATC 1482  
QY 420 ArgValThrAlaSerAspArgAspLYsGlySerAsnAlaValAlaHisTYrSerIleMet 439  
DB 1483 CGAGTCAAGCCTCGAGTCAAGCAAGGGAGCAATGCCGTGTGTGACATATAGCAATCAG 1542  
QY 440 SerGlyAsnAlaArgGlyGlnPheTYrLeuAspAlaGlnThrGlyValaLeuAspValVal 459  
DB 1543 AGTGGCAATCTCTGGGGAAGTTTATCTGAGATGCCCAAGTCACTGAGATGGG 1602  
QY 460 SerProLeuAspTYrGluThrTYrLYsGluTYrThrLeuArgValArgAlaGlnAspGly 479  
DB 1603 AGCCCTCTTGAATATAGACAGCAACAGAGTACACCTTACGGGTGAGAGACAGATGGT 1662  
QY 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499

DB 1663 GGCGCTCCCCCACTCTTAATGTCTCTGGTGGTGAAGATACAGGTCTGGATATCAAC 1722  
QY 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
DB 1723 GACAAATGCCCATCTTCTGTGAGCAACCCCTTCCAGGCTACTGTCTTGAAGAGCTCCCC 1782  
QY 520 LeuGlyTYrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
DB 1783 TTAGGCTACCTGTCTTCAATGTCCAGGCTATCCAGCTGATGTCTGTGACAAATCCCCG 1842  
QY 540 LeuGluTYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
DB 1843 CTGGAATACCCCTGTGCTGGGGTGGAGATGATCTTCCCTTCAACATCAAAATGGCACA 1902  
QY 560 GlyTPriIleSerValAlaIleGluLeuAspArgGluGluValaAspPheTYrSerPheLY 579  
DB 1903 GGCTGATCTCTGTGGCTGTGAACGTGACCGGAGGAAGTTGATTTTACAGCTTGGG 1962  
QY 580 ValGluAlaAspAsnHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
DB 1963 GTAGAGCTCGAGAACCAAGGACCTCCAGCACTCACTGCTGGCCAGTCAAGCTGACT 2022  
QY 600 ValLeuAspValaAsnAspAsnProThrPheThrGlnProGluTYrThrValArgLeu 619  
DB 2023 GTCTCGATGTCAACGACAAATCAACTTACCCCAACAGAGTACACAGTGGGCTC 2082  
QY 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValaAspArgAspAla 639  
DB 2083 AATGAGATGACGCTGTGGSCACAGCTGTGTGACGGTGAAGCTGTGGACCGGTGATCT 2142  
QY 640 HisSerValIleThrTYrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
DB 2143 CATAGTGCATCACTTACATCAATCAACAGTGGCAATTCGAAACCGCTTCCATCAACC 2202  
QY 660 SerGlnSerGlyGlyLYsLeuValSerLeuAlaLeuProLeuAspTYrLYsLeuGluArg 679  
DB 2203 AGCCAAAGTGTGTGGGCTGTATCCCTTGCCTGCCATCGAGCTAACAACTTGAAGCG 2262  
QY 680 GlnTYrValaValaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
DB 2263 CAGTATGTGTGGCTTTTACCGCTCCGATGGCACTCGGACAGGACCGGCAAGATGTG 2322  
QY 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYrThrVal 719  
DB 2323 GTGAATGCACCAAGCCCAACCAATCATGTCTGTCTTCAAGAGTCCCAATATCAAGT 2382  
QY 720 AsnValaAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
DB 2383 AATGTTAATGAGAACCGGCGGAGGACCAAGGTGTGTGATCAAGGCGCACGGAATGAG 2442  
QY 740 AspThrGlyGluAsnAlaArgIleThrTYrPheMetGluAspSerIleProGlnPheArg 759  
DB 2443 GACACAGTGAAGATGCCGATCACTTCACTTGAAGAGACAGCATCCCCCAATGCCG 2502  
QY 760 IleAspAlaAspThrGlyValaValThrThrGlnAlaGluLeuAspTYrGluAspGlnVal 779  
DB 2503 ATCGATGACAGACAGGGGGCTGTCAACCAAGCTGAGTGAATCAAGAGACCAAGTGT 2562  
QY 780 SerTYrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLYsSerAspThrThr 799  
DB 2563 TCTTACACCTCGCCATTAATGCTCTGGGACATGGCATTTCCCAAGAAATCCGACCAACC 2622  
QY 800 TYrLeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
DB 2623 TACCTGAGAGATCTGTGTGAACGATGAATGACATATGCCCTCAAGTCTCTCGAGACTCC 2682  
QY 820 TYrGlnGlySerValTYrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
DB 2683 TACCAAGGCAAGTGTATGAGAGATGTGCCACTTCACTACAGCTCTGCAATCTCAAGCC 2742  
QY 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTYrThrPheGlnGlyGlyAspAsp 859



Dh 2743 ACTGATCGTGAATTCGAGCTTAATGACGGGCTCTTCAACCTTCGACGAGGCGACGANT 2802  
Qy 860 G1AaPBG1YAaPPhel1eValG1uSerThrSerG1y1leValArgThrLeuArgLeu 879  
Dh 2803 GGAAGCGGTGACTTATTGTTGAGTCCAGTCAAGGCACTGTCGACGACGCTACGAGGCTG 2862  
Qy 880 AaPArgG1uaSerVal1aG1nTrVal1leuArg1a1Tyra1aValAaP1yG1uMetPro 899  
Dh 2863 GATTCAGAGAAAGTGGCCCAAGTATGTCTTGCGGGCATATGCAGTGACAAAGGAGATGCCC 2922  
Qy 900 Pro11aArgThrProMetG1uVal1ThrVal1ThrVal1leuAaPValAaAaPAsnProPro 919  
Dh 2923 CAGGCCCGACACCTATGAGAAATGACAGTCACTGTGTGATGATGAAATGACATCCCCCT 2982  
Qy 920 Val1PheG1uG1naAaP1uPheAaPVal1PheVal1G1uG1uaSerPro11eG1y1leu1a 939  
Dh 2983 GTCCTTGAGCAGAGAGAGTTGATGTGTTTGAGAAAGAAACGCCCATTTGGGCTTAAGCC 3042  
Qy 940 Val1AaArgVal1ThrAlaThrAaPProAaP1uG1yThraAaA1aG1n1leuMetTyG1n 959  
Dh 3043 GTGGCCCGGTGTCACAGCACTGACCCCGATGAAAGGCAACAATGCCCAGATTATGTACAG 3102  
Qy 960 11eValG1uG1yAaA11eProG1uVal1PheG1uLeuAaP11ePheSerG1yG1uLeuThr 979  
Dh 3103 ATGTGAGAGGCAACATCCCTGAGGTCCTTCAAGTGAACATCTTCCGGGAGCTGACA 3162  
Qy 980 A1aLeuValAaPLeuAaP1yG1uAaAaPArgProG1uTyra1leuVal11eG1nA1aThr 999  
Dh 3163 GCCCTGAGACTTATGACTACAGAGACCGGCTGAGTACGTCCTGTATTCAGGCCACG 3222  
Qy 1000 Ser1aAProLeuVal1SerArg1a1ThrVal1HiSVal1ArgLeuLeuAaPArgAaAaPAsn 1019  
Dh 3223 TCAGCTCCTGTGTGAGCCGGGCTACAGTCCACGTCGCTCTTGAACCGCAATGACAAC 3282  
Qy 1020 ProProVal1leuG1yAaA11ePheAaPAsn1y1leuPheAaAaP1yVal1ThraAaArgSer 1039  
Dh 3283 CCACCAAGTGTGGGCACTTGTAGATCTTTCAACAACATATGCAACATCGCTCAAGC 3342  
Qy 1040 SerPheProG1yG1yA1a11eG1yArgVal1Pro1aHiSAsPProAaP11eSerAaPser 1059  
Dh 3343 AGCTTCCCTGGGGGGTCCATTCGCGAGTACCTGCCCATGACCTCGATATCTCAGATAGT 3402  
Qy 1060 LeuThrTy1ySerPheG1uArgG1yAaA11eAaP1uSerLeuVal1leuAaA1aSerThr 1079  
Dh 3403 CTGACTTACAGCTTGAAGCGGGGAATGACCTGAGCTGCTCAATGCTCCACG 3462  
Qy 1080 G1yG1uLeuTy1yLeuSerArg1a1leuAaAaPAsnAaPProLeuG1uA1a11eMetSer 1099  
Dh 3463 GGTAGCTGAAGCTTAAGCCGCGCACTGGAACAACCGGCTCTGGAGGCAATCAATGAGC 3522  
Qy 1100 Val1leuVal1SerAaP1yVal1HiS1eSerVal1Thra1aG1nCySA1aLeuArgVal1Thr11e 1119  
Dh 3523 GTGCTGTGTGACAGCGGCTACAGCGCTGACCGCCAGTGGCGCTGCTGTACATC 3582  
Qy 1120 11eThraAaP1uMetLeuThrHiS1eSer11eThraLeuArgLeuG1uAaPmetSerProG1u 1139  
Dh 3583 ATCAACCAATGATGCTCAACCAAGCATCAAGCTGCGCTGAGGAGCATGTCAACCCAG 3642  
Qy 1140 ArgPheLeuSerProLeuLeuG1y1leuPhe11eG1nA1aVal1A1a1leuAaA1aThr 1159  
Dh 3643 CGCTTCTGTGACCACTGATAGGCTCTTCACTCAAGCGGAGGCGCCACGCTGGCCAG 3702  
Qy 1160 ProProAaP1a1aVal1a1aPheAaA1aG1naAaP1yThraAaP1aA1aProG1yG1yHiS 1179  
Dh 3703 CCACCGAGCAAGTGTGTGCTTCAAGTACAGCGGCAACCGAGCCCGGGGGGCGCAC 3762  
Qy 1180 11eLeuAaA1aSerLeuSerVal1G1yG1nProProG1yProG1yG1yProProPhe 1199  
Dh 3763 ATCTCAACGTAAGCTGTGTGTGGCCAGCGGCAAGGCGCGGGGGCGGCGCTTTC 3822  
Qy 1200 LeuProSerG1uAaPLeuG1nG1uArgLeuTy1yLeuAaAaPserLeuThrAla11e 1219  
Dh 3823 CTGCTCTGTAGAGACTGACGAGGCGCTTATCTCAACCGGAGCTTGTACGCGCATC 3882

Qy 1220 Ser1aG1naArgVal1leuProProPheAaPAsn11eCyLeuAaArg1uProCyG1uAa 1239  
Dh 3883 TCGGCAAGCGCGTGTGCTTCCATGACAAACATCTGCTCGGAGACCTTGGAGAAC 3942  
Qy 1240 TyraMetArgCySA1eSerVal1leuArgPheAaPserSer1aProPhe11eA1aSerSer 1259  
Dh 3943 TACATGGCGCTGCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4002  
Qy 1260 SerVal1leuPheAaPPro11eHiSProVal1G1yG1uAaArgCySA1yCySA1yProProG1y 1279  
Dh 4003 TCGTGTCTTCCGCGCCATTCACCCGCGAGAGGCTGCGCTGCGCTGCGCTGCGCTGCG 4062  
Qy 1280 PheThrG1yAaP1yTyraG1uThra1aValAaPLeuCySA1yTyraArgProCyG1yPro 1299  
Dh 4063 TTCAACGGTGACTATCTGAGAGACCGAAGTGAACCTTCTGCTGCTGCTGCTGCTGCTG 4122  
Qy 1300 HiG1yArgCySA1yAaPserArgG1yG1yTyraThrCySA1yCySA1yArgAaP1yTyraThr 1319  
Dh 4123 CACGGGCGCTGCGCAGACCGGAGGCGGCTACACCTGCTGTGATGTGCTACACG 4182  
Qy 1320 G1yG1uHiS1yG1uVal1Ser1a1aArgSerG1yArgCySA1yProG1yVal1CySA1yAa 1339  
Dh 4183 GGTAGGACGTGAGGTGAGTGTGCTGCTGAGCGCGTTCACCCGGGTCTGCAAGAT 4242  
Qy 1340 G1yG1yThraCySA1aAaPLeuLeuVal1G1yG1yPhe1yCySA1yAaP1yProSerG1yAaP 1359  
Dh 4243 GGGGAGACCTGTGCAACCTGCTGTGGGGGTTTCAAGTGCATTCGCCATCTGGAGAC 4302  
Qy 1360 PheG1u1yProTyraG1nVal1ThrThraArgSerPhePro1aHiS1ePhe11eThr 1379  
Dh 4303 TTCCAGAAAGCCCTATCTCCAGTGAACACGCGAGCTTCCCGGCCACTCCTTCATCAC 4362  
Qy 1380 PheArgG1y1yLeuArgG1naArgPhe11ePheThra1a1eAaPserPhe1a1aThrTyG1u 1399  
Dh 4363 TTTCGCGCGCTGCGCAGCGGTTCACCTTCACTCACTGCGCTCTGTGTCACAAAGAG 4422  
Qy 1400 ArgAaP1yLeuLeuLeuTyraAaG1yArgPheAaG1yTyraAaP1yAaP1yAaP1yAaP1y 1419  
Dh 4423 CGGACGCGGTGTGCTGTGTCATGAGCGCGTTTCAATGAAGACATGACTTTGGGCCCTC 4482  
Qy 1420 G1uVal11eG1nG1uG1nVal1G1nLeuThrPheSer1aG1yG1uSerThrThrVal1 1439  
Dh 4483 GAGGTGATCCAGAGGAGGTGCAAGTCACTTCTGCAAGGAGTCAACCAACCGGTG 4542  
Qy 1440 SerProPheVal1ProG1yG1yVal1SerAaP1yG1yThra1a1eAaP1yThra1a1eAaP1y 1459  
Dh 4543 TCCCATTTGTCGCGGAGGAGTCAAGTATGCGCATACGTGCAAGCTGAATAAC 4602  
Qy 1460 TyraAaP1yAaPLeuLeuG1yG1nThra1yLeuProG1nG1yProSerG1uG1nTyraVal1 1479  
Dh 4603 TACAAATAGCACTGTTGGGTACAGAGGCTCCCAAGGCGCCATGACAGAGAAAGTGTG 4662  
Qy 1480 A1aVal1a1ThrVal1AaP1yCySA1yThra1yVal1A1aAaP1yPheG1ySerVal1leu 1499  
Dh 4663 GCTGTGTGACCGGTGATGCTGTGACACAGAGTGTGCTTGGAGTGTGCTGCTGCTG 4722  
Qy 1500 G1yAaA1ySerCySA1a1aG1nG1yThra1G1yG1ySerTyra1ySerLeuAaPLeu 1519  
Dh 4723 GGCACTATCTGTGTGCTGCGCAAGGCAACCAAGGAGTGAAGAAAGTCTGTGAATCTG 4782  
Qy 1520 ThraG1yProLeuLeuLeuG1yG1yVal1ProAaPLeuProG1uSerPheProVal1ArgMet 1539  
Dh 4783 ACGGAGCGCTGTGCTAAGCGGGGTGCTGACCTGCGAGAGCTTCCAGTCCGAATG 4842  
Qy 1540 ArgG1nPheVal1G1yCySA1eAaAaPLeuG1nVal1AaPserArgHiS11eAaPmetA1a 1559  
Dh 4843 CCGGAGTGTGTGAGCTGCAAGCGGAACCTGCAAGTGAACCGGCGCAATAGCATGTGCT 4902  
Qy 1560 AaPPhel1eA1aAaAaG1yThra1a1ProG1yCyProA1a1yAaP1yAaA1a1e 1579  
Dh 4903 GACTTATTCGAACAAATGACACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4962

QY 1580 SerAenThrCyHhSaenGlyYThrCyVaSaenGlnTPaPaLaPheSerCyGlu 1599  
 Db 4963 AACCAACACTTGCACATGGGGGCACTTGCATGAAACATGGGAGCGCTTACGCTGCGAG 5022  
 QY 1600 CysPProLeuGlyPheGlyGlyYSerSerCySaIaGlnGlnMetAlaAsnProGlnHsPhe 1619  
 Db 5023 TGCCTCCCTGGGCTTGGGGGCAAGAGCTCGCCGCAAGAAATGGCCATCCACAGCACTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTrpHsGlyLeuSerLeuProIleSerGlnProTrpYr 1639  
 Db 5083 CTGGGAGAGAGCTGGTGGCTGGCAATGGCTCTGCTCCATCTCCCAACCTGGATGC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTCAGCTCATGCTTCCGACAGCGCCACAGCCGAGGCTGCTGCTGACGCGCATTCACAG 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyValMetLeuSerValGlnGly 1679  
 Db 5203 GGGCGGACGACCTTACCTTACAGCTACAGAGAGGCGACGTATGCTGAGCGTGAAGGCG 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrp 1699  
 Db 5263 ACAGGGCTTCAGGGCTCTCTCTCCGCTGAGGACAGCGCGGCGCATACGCTGACTGG 5322  
 QY 1700 HsHsAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyValAlaIleLeuSerPhe 1719  
 Db 5323 CACCATGCAAGCTGGCACTGGGAGCGCAGCGGGGGCTTGGCCATGCTCATTCGTCCTTC 5382  
 QY 1720 AspTYrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHsGlyLeuHsGln 1739  
 Db 5383 GATTATGGGACAGAGAGAGAGGAGGACCTGGGCGCCCGGCTGATGCTGACCTGG 5442  
 QY 1740 SerAsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaIaArgIlePheArg 1759  
 Db 5443 AACCAACATPACAGTGGGGCGGAATACCTGGGCGAGCGGGGTGGCTGGCTTGGG 5502  
 QY 1760 GlyCySLeuGlnGlyValArgValSerAspThrProGlnGlyValaIaAsnSerLeuAspPro 1779  
 Db 5503 GCGCTGTTTCAGAGGTGTGGGTGAGCAGATACGCCAGAGGGGTGTAAACGCTGATGCC 5562  
 QY 1780 SerHsGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSer 1799  
 Db 5563 AACCATGGGAGAGCATCAACGTGAGCAAGGCTGTAAGCTGCTCAACCTTGTATCTCA 5622  
 QY 1800 AsnProCysProAlaAsnSerTYrCysSerAsnAspTrpAspSerTYrSerCysSerCys 1819  
 Db 5623 AACCCGTGCTGCTTACAGCATATTCAGCAACGATCGGGAACGATATTCGAGACTGT 5682  
 QY 1820 AspProGlyTYrTYrGlyArgAsnCySThrAsnValCysAspLeuAsnProCySLeuHs 1839  
 Db 5683 GATCCAGGTTTACTATGGTGAACACTGATATGTGTGACCTGTAACCCGTGTAGCAC 5742  
 QY 1840 GlnSerValCysThrArgIysProSerAlaProHsGlyTYrThrCysGlnCysProPro 1859  
 Db 5743 CAGTCTGTGTGACCCGCAAGGCCAGTGGCCCCCAATGCTATACCTGCAAGTGTCCCCA 5802  
 QY 1860 AsnTYrLeuGlyProTYrCysGlnThrArgIleAspGlnProCysAspProArgIleTYr 1879  
 Db 5803 AATTACTTGGGCGCATACTGTGAGACCAAGATTGACCAAGCTTGTCCCGGTGGCTGTGG 5862  
 QY 1880 GlyHsProThrCysGlyProCysAsnCySAspValSerIleGlyPheAspProAspCys 1899  
 Db 5863 GAGCATTCACCAAGTGGCCATGCAACTGTGATGACGAAGAGCTTTTACCCAGACTGC 5922  
 QY 1900 AsnIleThrSerGlyGlyCysHsCySlyGlnAsnHsTYrArgProProGlySerPro 1919  
 Db 5923 AACAAACAGAGGGCGAGTGCCTGCAAGAGAACATACCGGCGCCCAAGGCAAGCCCC 5982  
 QY 1920 ThrCysLeuLeuCysAspCysTYrProThrGlySerLeuSerArgValCysAspProGlu 1939  
 Db 5983 ACCTGCTCTGTGTGACTGCTACCCACAGGCTCTGTGCTGAGACTGTGACCTGTAG 6042  
 QY 1940 AspGlyGlnCysProCysIysProGlyValIleGlyArgGlnCysAspArgCysAspAsn 1959

Db 6043 GATGGCAGTGTTCATGCAACGAGTGTATCGGGCGTCAGTGTGACCGCTGTGACAC 6102  
 QY 1960 ProPheAlaGlyValAlaThrThrArgGlyCysGlu-----ValAsnTYrAspSer 1975  
 Db 6103 CTTTGTGCAAGTGCACCAATGGCTGTGAAGGGCCCTTGTGTGATTAAC----- 6156  
 QY 1976 CysProAlaAlaIleGlnAlaGlyIleTYrTrpProArgThrArg----- 1990  
 Db 6157 TGTCCCGGCCCATAGAGTGC-----TGGCTCCAGAGAAACCTTCAGCCAGCT 6207  
 QY 1991 PheGlyLeuProAlaAlaIleProCysProIysGlySerPhe-----Gly 2005  
 Db 6208 CAGGGGCTTCCGTGTGTCTCTCT-----GAGGCGGCGCTTGTGGCTTCCCCCAAGG 6264  
 QY 2006 ThrAlaValArgHsAcCysAspGlnHsArgGlyTYrLeuProProAsnLeuPheAsnCys 2025  
 Db 6265 ACTGCTGTGGCAGCTGTATGAGCAGAGGGGGTGGCTCCCCCAAACTCTTCAACTGC 6324  
 QY 2026 ThrSerIleThrPheSerGlnLeuLeuYsGlyPheAlaGlyLeuGlnArgAsnGlySer 2045  
 Db 6325 ACGTCATCACTTCTCAGAACTGAAGGGCTTCGCTGAGCGGCTTACAGCGGAATGAGTCA 6384  
 QY 2046 GlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuAsnAlaThrGlnHs 2065  
 Db 6385 GGCCTAGACTCAGGGCGCTCCACAGCAGTACCGCTGCTCGCGCAACGCCACGACGACAC 6444  
 QY 2066 ThrAlaGlyTYrPheGlySerAspValIysValAlaTYrGlnLeuAlaThrArgLeuLeu 2085  
 Db 6445 ACGCTGAGCTACTTCCGAGGAGAGGTCAAGGTGGCTTACAGCTGGGCAACCGGCTGCTG 6504  
 QY 2086 AlaHsGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHsPheThr 2105  
 Db 6505 GCCCAACAGAGACACCAAGCGGGGCTTGGCTGTCTGCCACACAGGACGTGCACTTCACT 6564  
 QY 2106 GlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsnIysArgHsIleProGlu 2125  
 Db 6565 GAGAAATGCTGGCGGTGGGAGCGGCTCTGTGACACAGCAACAAACGGGCACTGGGAG 6624  
 QY 2126 LeuIleGlnGlnThrGlnGlyGlyTYrThrAlaTrpLeuLeuGlnHsIleTYrGlnAlaIleTYrAla 2145  
 Db 6625 CTGATCCAGAGACAGAGGGGTGGACCGCTGGCTGCTCAGCACTATGAGGCGTACGCC 6684  
 QY 2146 SerAlaLeuAlaGlnAsnMetArgHsIleThrTYrLeuSerProPheThrIleValaIlePro 2165  
 Db 6685 AGTGCCCTGGGCCAGAAATCGGCACTACCTAACCCCTTCAACATGCTCACGGCC 6744  
 QY 2166 AsnIleValIleSerValValArgLeuAspIysGlyIysAsnPheAlaGlyValaIleLeuPro 2185  
 Db 6745 AACATTTGATCTCCGTAGTGGCTTGGACAAAGGAACTTGTGCTGGGGCCAAAGCTGCC 6804  
 QY 2186 ArgTYrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThrThrValIleLeuPro 2205  
 Db 6805 CCGTACAGAGGCGCTCGTGGGAGAGCGCCCGGACCTTGAAGACAAACGATCTGTGCT 6864  
 QY 2206 GlnSerValPheArgGlyThrProProValValArgProAlaGlyProGlyGlnAlaGln 2225  
 Db 6865 GAGTCTGTCTTCAAGAGAACCGCCCGGTGTGAGGCCCGGAGGCGCCGAGAGGCGCAG 6924  
 QY 2226 GlnProGlnGlnLeuAlaArgArgGlnArgHsPProGlnLeuSerGlnGlyGlnAla 2245  
 Db 6925 GAGCCAGAGAGCTGGCAAGGAGACAGGAGCGCACCCGAGAGCTGAGCCAGGAGTAAAGCT 6984  
 QY 2246 ValAlaSerValIleIleTYrArgThrLeuAlaGlyLeuLeuProHsAsnTYrAspPro 2265  
 Db 6985 GTGGCAAGCTCATATCACTACCGACCTGGGCGGGCTTACTGCTCATATGATGACCT 7044  
 QY 2266 AspIysAspSerLeuArgValProIysArgProIleIleAsnThrProValIaSerIle 2285  
 Db 7045 GACAAAGCCAGCTTGAAGTCCCAAGCGCCGATCAACACACCGGTGTGAGCATC 7104  
 QY 2286 SerValHsAspAspGlnGlnLeuLeuProArgAlaLeuAspIysProValaIleGln 2305

Db 7105 AGCTTCATGATGATGAGGAGCTTCTGCCCCGGGCGCTGACAAACCCCTGACGGTGACG 7164  
 Qy 2306 PheATgLeuEuglUThrGluGluArgThyLysProIleCysValPheThrPheHisSer 2325  
 Db 7165 TTCGGCGCTGCGAGACAGAGAGCGGACCAAGCCATCTGTCTTCTGGAACCATTTCA 7224  
 Qy 2336 IleLeuValSerGlyThrGlyGlyTyrSerAlaIleArgIleCysGluValValPheArgPhe 2345  
 Db 7225 ATCTCGGTCACGTGGGACAGGTGCTGTGTGGCCAGAGGCTGTGAAGTCTCTTCGCAAT 7284  
 Qy 2346 GluSerHisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspVal 2365  
 Db 7285 GAGAGCCACGATGAGCTGCGAGTGCACACATGACGAGCTTGTGTGTCTCAATGACGTT 7344  
 Qy 2366 SerATgATgGlu-----AsnGlyGluIleLeu 2374  
 Db 7345 TCTCGCGCGGAGGCTGGGCGCCACAGGGGCGAGCTGCAGAGCCGTGGAATGGGAGATCTCTG 7404  
 Qy 2375 ProLeuLysThrLeuThyTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThr 2394  
 Db 7405 CCACCTGAAAGACCTGACATACCTGTGCTTGAAGTGTACCTGTGGCTGCTCTGCTCAC 7464  
 Qy 2395 PhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyTleArgArgPhe 2414  
 Db 7465 TTCTTCTTCTTCACTCTTGTGGTATCTGTGCTCCACACCAACGAGCATCCGAGCTTAC 7524  
 Qy 2415 LeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAsp 2434  
 Db 7525 CTGACAGCTGGCGCTGGGCGCTGAGCTGTCTTCTCTGGGATCAACAGGCTTAC 7584  
 Qy 2435 LeuProPheAlaCysThyValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPhe 2454  
 Db 7585 CTCCCTTGTGCTGACAGTCACTGCACTCCGCTGCACTTCTTACCTCTGCACTTT 7644  
 Qy 2455 SerTrpAlaLeuLeuGluAlaLeuHisIleLeuTyrArgAlaLeuThyGluValArgAspVal 2474  
 Db 7645 TCTTGCGCTTGTCTGAGGCTTGTGACCTGTACCGGGGCACTACAGAGGTGGCCATGTC 7704  
 Qy 2475 AsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr 2494  
 Db 7705 AACACCGGCGCCCATGCGCTTCTACATACGCTGGGCGTGGGCGTGGCTTCACTCA 7764  
 Qy 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSer 2514  
 Db 7765 GGGCTAGCGCGTGGGCGCTGACCCCGAGGGCTACGAGAAACCTGACTTCTGCGCTCTCC 7824  
 Qy 2515 IleTyrAspTrpLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSer 2534  
 Db 7825 ATCTATGACAGCTCATCTGAGATTGTGTGGCCCGGTGGCTTGTGCGTCTCGATGAGT 7884  
 Qy 2535 ValPheLeuTyrIleLeuAlaIleArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlu 2554  
 Db 7885 GTCTTCTGTATCATCTGGCGGCGCGGCGCTCTGTGTGGCCAGCGGCAAGGCTTTAG 7944  
 Qy 2555 LysLysGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAla 2574  
 Db 7945 AAGAAAGGTCTGTCTGCGGCTGACGCGCTCTTGTGCGCTCTCTCTGCTGAGCGCC 8004  
 Qy 2575 ThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspTrpLeuPheHisTyrLeuPhe 2594  
 Db 8005 ACGTGCGCTGCGACCTGCTCTCTTCAACAGCAACCTCTCTTCACTCACTCTT 8064  
 Qy 2595 AlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLys 2614  
 Db 8065 GCTACTCTGCAATGATCAAGGCGCCCTTCACTCTCTCTATGTGTGTCTTAGCAAG 8124  
 Qy 2615 GluValArgLysAlaLeuLysLeuAlaCysSerArgLysProSerProAspProAlaLeu 2634  
 Db 8125 GAGGCTCGGAAAGCACTCAAGCTTCTGCAAGCCGCAAGCCCGCAAGCCCTGCTGTG 8184  
 Qy 2635 ThrThyLysSerThrLeuThySerSerTyrAsnCysProSerProTyrAlaAspGlyArg 2654  
 Db 8185 ACGACCAAGTCAACCTGACCTGCTCTACACTGCCCCAGCCCTTACGCAAGATGGCGG 8244

Qy 2655 LeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLys 2674  
 Db 8245 CTGTACAGCCCTTACGAGGACTCGGCGGCTCTTGTGCAACAGCAAGTGTGCTGGGCAAG 8304  
 Qy 2675 SerGlnProSerTyrIleProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyGln 2694  
 Db 8305 AGTCAGCCCACTCATCCCTTCTGTGTGAGGGAGAGTCCGCACTGAACCTTGGCCAA 8364  
 Qy 2695 GlyProProGlyLeuGlyAspProGlySerLeuPheLeuGlnGlnGlnHis 2714  
 Db 8365 GGGCCCCCTGGCGTGGGGGATCCAGGCGCTGTCTTCTGGAAGTCAAGACAGACGAT 8424  
 Qy 2715 AspProAspThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerGlySerTyr 2734  
 Db 8425 GATCTGACACAGCACTCCGACAGTACTTCTTGAAGACGACAGATGTGCTCTTAT 8484  
 Qy 2735 AlaSerThrHisSerSerAspSerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2754  
 Db 8485 GCTCTACCCCACTCATCAGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8544  
 Qy 2755 PheProGlyGlyGlnGlyTyrTrpAspSerLeuLeuGlyProGlyAlaArgLeuProLeu 2774  
 Db 8545 TTCCCTGAGAGACAGGCTGGGATAGCTGTGTGGGCGCTGAGAGACAGACATGCTGCTG 8604  
 Qy 2775 HisSerThrProLysAspGlyGlyProGlyProGlyLysAlaProTyrProGlyAspPhe 2794  
 Db 8605 CACAGTACTCCCAAGAGATGGGGGCGCCAGAGGCTGTGGCAAGGCCCCCTGGCCAGAGACTTT 8664  
 Qy 2795 GlyThrThrAlaLysGlySerSerGlyAsnGlyValaProGlnGlnArgLeuArgGluAsn 2814  
 Db 8665 GGGACCAACAGAAAGAGAGATGAGGCAACGGGGCGCCCTGAGAGAGCGGCTCGGAGAAAT 8724  
 Qy 2815 GlyAspAlaLeuSerArgGlyGlySerLeuGlyProLeuProGlyLysSerSerAlaGlnPro 2834  
 Db 8725 GGAATGCGCTGTCTCAGAGGGGTCTCTAGGCCCTTCTTCAAGGCTCTTGTGCCAGGCT 8784  
 Qy 2835 HisLysGlyIleLeuLysLysCysLeuProThrIleSerGlyLysSerSerLeuLeu 2854  
 Db 8785 CACAAAGGCATCTTAAAGAGAGATGTCTGCGCACATACAGAGAGAGAGAGAGAGAG 8844  
 Qy 2855 ArgLeuProLeuGlnGlnCysThrGlySerSerArgLysSerSerAlaSerGlyLysSer 2874  
 Db 8845 CGGCTCCCCCTGAGCAATGACAGGGGTCTTCCCGGGGCTCTCCGCTAGTGAAGGGCAGC 8904  
 Qy 2875 ArgLysGlyProProProAspArgProProProArgGlnSerLeuGlnGlnLeuAsnGly 2894  
 Db 8905 CGGGGCGGCGCCCTTCCCGCCACCGCGCGGAGAGCTTCCAGAGAGAGAGAGAGAGAG 8964  
 Qy 2895 ValMetProIleAlaMetSerIleLysAlaGlyThrValAspGluAspSerSerGlySer 2914  
 Db 8965 GTCATGCGCAATCGCAAGAGATCAAGGAGGAGGAGGAGGATGAGACTGCTCAAGGCTCC 9024  
 Qy 2915 Glu 2915  
 Db 9025 GAA 9027

RESULT 6  
 LOCUS AB028499 8950 bp mRNA linear ROD 06-SEP-1999  
 DEFINITION Mus musculus mRNA for Flamingo 1, complete cde.  
 ACCESSION AB028499  
 VERSION AB028499.1 GI:5832710  
 KEYWORDS Flamingo 1.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (slices)  
 Ueul, T., Shima, Y., Shimada, Y., Hirano, S., Burgess, R. W.,  
 Schwartz, T. L., Takeichi, M. and Uemura, T.  
 Flamingo, a seven-pass transmembrane cadherin, regulates planar  
 cell polarity under the control of Fz/zed



Db 841 GACACCAACGATCAGACCCCTTTTGAACGCAAGATACAGAGAGCCCTCAGGGAG 900  
 QY AsnLeuGluValGlyTyrGluValLeuThrValAlaGlyAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGAGAGGTTGGCTATGAGGTGCTTACAGTCAAGGCGCACCGAGCGCCCTTCCC 960  
 QY AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyGlySerProSerGlyValAlaPhe 340  
 Db 961 AATGCAACATTCGTGTACCGCTGCTGAGGGGGGAGAGACAGCCCTCAGACCCCTTT 1020  
 QY GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 1021 GAGATGATCTCTCGTCCGGGGGTGATCCGAAACCCCGGCTCTGAGACCGGGAGAGATG 1080  
 QY GluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1081 GAATCCTTCAAGATTAAAGTGGAGCGAGTACAGAGGCGCGGAGCCAGGCCCGGAGGT 1140  
 QY ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 Db 1141 TCCACAGCCATGTGTTCTCTGCTGAGAGATGATTAATGACAAAGCCCCAGTTTAC 1200  
 QY GluIleAspArgTyrValValGluValAlaArgIleAspValThrProGlyAlaProValLeuArg 420  
 Db 1201 GAGAAAGCTTATGTGTCTCAGGTGCGGAGAGTGTGACCCAGAGCCCTCAGTTCTCGA 1260  
 QY ValThrAlaSerAspArgAspIleGlySerAsnAlaValAlaIleTyrSerIleMetSer 440  
 Db 1261 GTCAACCGCTCCGANTAGAGACAGAGGCGAGCATGCTTAAGTGCATACAGATCAAGAT 1320  
 QY GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValAlaSer 460  
 Db 1321 GGCATGTCTGGGGGCGAGTTCTATCTGATGCTCAGCTGAGCCCTGAGTGTGATAGC 1380  
 QY ProLeuAspTyrGluThrThrLeuGluTyrThrLeuArgValArgAlaGlyAspGly 480  
 Db 1381 CCACTCCAGCTATGAGACAAACAAAGATTAACGCTTCGAGATCCGGAGCTCAGAGCGTGC 1440  
 QY ArgProProLeuSerAsnValSerGlyLeuValThrValGluValLeuAspIleAsnAsp 500  
 Db 1441 CGTCCCTCAGCTCCAAATGTCTCTGGCTTAGTAACAGTGCAGAGTCTTAGATCAAGAT 1500  
 QY AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
 Db 1501 ATACGGCCCCCATCTTGTGAGCAGCCCTTTCAGGCGCACCGTCTAGAGAGTGTCTCT 1560  
 QY LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 Db 1561 TTAGGCTACCTTGTCTTCTGACGTCAGGCGCAATTGATGCTGAGCTGTGATTAACGCCGC 1620  
 QY LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
 Db 1621 CTCGAGTATACCTCGGAGGGGTTGGGCAAGACTTCTTCCATTAACAATGAGCA 1680  
 QY GlyTyrPheSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 Db 1681 GGCAGGATCTGTGGCGGAGAGTGGATCGGAAAGGTTGATTCTAAGCTTTGGT 1740  
 QY ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 Db 1741 GTAAAGGCCCGGAGCAATGAGCAACCCAGGCTCAGGCTCAAGCAAGTCAAGGTAAAC 1800  
 QY ValLeuAspValAlaAspAspAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
 Db 1801 ATCTTGAGCGTCAATGACAAACCACTTACGAGCGAGGTACACCGTTCCGCTC 1860  
 QY AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAla 639  
 Db 1861 AACAGAGAGCGCGCGGAGCAAGTGTGTGACCGTCTCAGCGGTGAGTCAAGCAAGCT 1920  
 QY HisSerValIleThrTyrGluIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 Db 1921 CAGAGTGCATTAACCTACCAATCAACAGCGGCAACCGGCAACCGCTTCTATAC 1980

QY SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluAlaArg 679  
 Db 1981 AGCCAAGTGTGTGATCTGTCTCTTCCCTTACCGCTAGACTACAACTGAGCGG 2040  
 QY GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgIleAspThrAlaGlnIleVal 699  
 Db 2041 CAGTATGTGTGGCAGTGAAGTGCCTCAGATGCGACAGGCGAGCACAGCTCAATGATA 2100  
 QY ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
 Db 2101 GTGATGTCAACCGATGCTCAACCCAGCTGCTCCGCTTCCAGACCTCCACTACAGATC 2160  
 QY AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2161 AACGTATATGAGACCGGCGAGAGGAGCACAGTGTGCTGATCAAGCCGACGATGAG 2220  
 QY AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2221 GACACGGGGAGATGCCCCGATCACTTATGAGATGATGATCCCCAGTTCCGA 2280  
 QY IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 Db 2281 ATCGACGGGAGCACTGGGGCTGTCAACCCAGGCTGAGCTGAGCTACAGAGACAGGTG 2340  
 QY SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
 Db 2341 TCTTATACCTGTGGCAGTCACTGCTCGGAGCAATGCGATTTCCAGAGTGTGACACACC 2400  
 QY TyrLeuGluIleLeuValAlaAspValAlaAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 2401 TATTGAGATGCTCTGGTGAATGACGTGACACACCCCGCTTCTCAGAGATTCC 2460  
 QY TyrGlnIleSerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 2461 TACAGGGGACCGCTACAGAGAGAGTGCCTTCCAGCGCTGCGAGATCTTAGCC 2520  
 QY ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnIleGlyAspAsp 859  
 Db 2521 ACGGATGAGATCTCGGCTCGAATGGCAGGGTTTCTACACCTTTCAGAGAGAGATGAT 2580  
 QY GlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLeu 879  
 Db 2581 GGGGACGTGATCTTATTTGATGAGTGCACATCGGGCATTTGTGGCACATCGCGGAGCTG 2640  
 QY AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIleGlyMetPro 899  
 Db 2641 GATGTGAGAACGTGGCCAGTACGTTCTGGCGGCTCATCGGAGTGCAGAGGCAATGCCA 2700  
 QY ProAlaArgThrProMetGluValThrValThrValLeuAspValAlaAspAsnProPro 919  
 Db 2701 CAGGCCGAGCAACCAATGAGAGTGAAGTACTGTCTGAGATGGAGATGACAAACCCCT 2760  
 QY ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 2761 GTCTTGAACAGATGATGTTGACGTATTGTGAGAGAGAACGCCCATCGGGCTGGCC 2820  
 QY ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 2821 GTGGCCGGGATCAGGCACTGACCGAGTGAAGAGAGCAACGACAGATCAATGATACAG 2880  
 QY IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 2881 ATTGTGAGGGCAATATCCCTGAGGCTTCACTTCACTGATGATATATCTTCTGGCGAGTACT 2940  
 QY AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 2941 GCCCTGTATATTGACATGAGAGACCGGCTGAATATGTTCTGTGATCAACCGGCTACG 3000  
 QY SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3001 TCTGCTCCCTTGTGAGAGGCTACTGTCCATGTCCGCTCTTACCGCATGATAC 3060



QY 1020 ProProValLeuGIYAsnPhelGluIleuPheAsnIleuValThrAsnArgSerSer 1039  
DB 3061 CGCCCGGCTGCGGCACTTTAGATCTCTTACACATATGTCACCAACCGCTCCACC 3120  
QY 1040 SerPheProGIYGIYAlaIleGIYArgValProAlaHisAspProAspIleSerAspSer 1059  
DB 3121 ACCTTCCCGGCGGTCGATAGGCGGTGCTGCCACGACCCCGATATCTCAGACAGC 3180  
QY 1060 LeuThrTrpSerPheGluIleuValGlnIleuSerLeuValLeuLeuAsnAlaSerThr 1079  
DB 3181 CTGACGTATACGCTGAGGAGGAGAAATGAATCACTGCTGCTGCTCAATGCTCCACC 3240  
QY 1080 GIYGIYLeuLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGIYAlaIleMetSer 1099  
DB 3241 GGTGAGCTGAGACTGAGCCGGGCACTGACACACACCGGCCCTTGAGACCTATGAGT 3300  
QY 1100 ValLeuValSerAspGIYValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
DB 3301 GTGCTGCTGTCAGATGCTGTCACAGTGTGACAGCCCACTGCTGCTGCTCAGCANTC 3360  
QY 1120 IleThrAspGluMetLeuThrHisSerIleThrIleuArgLeuGluAspMetSerProGlu 1139  
DB 3361 ATCAAGAGAGATGCTCAACACAGCATACGCTGCTGAGAGACATGTTCTCCAGAA 3420  
QY 1140 ArgPheLeuSerProLeuLeuGIYLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
DB 3421 CGCTTCTCTCAACCTGCTGGGACCTTCACTTACGCGCTGGGCACTTGGCCACA 3480  
QY 1160 ProProAspHisValValAlaPheAsnValGlnArgAspThrAspAlaProGIYGIYHis 1179  
DB 3481 CCCCAGGATCATGTGTGCTCTTCAATGTGCAAGAGGATATGATGCCCAAGCGCCCAT 3540  
QY 1180 IleLeuAsnValSerLeuSerValGIYGIYProProGIYProGIYGIYProProPhe 1199  
DB 3541 ATCTTCACGTAAAGCTGTGCTGAGGCGCAGCCCTCCAGACCCGGGGGTGGCCACCTTT 3600  
QY 1200 LeuProSerGIYAspLeuGlnIleuArgLeuIYLeuAsnArgSerLeuLeuThrAlaIle 1219  
DB 3601 CTGCTTCAAGAGATCTCCAGAGGCGCTGTACTCAACGAGCGCTGCTCAACCGGCANTC 3660  
QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAs 1239  
DB 3661 TCGGCGAAGCGGTGCTCCCTTCGACACACAACTTGTCTG-CGGAGACCTTGGAGAA 3719  
QY 1239 nIYrHeArGYSValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSe 1259  
DB 3720 TTACATGCAATGTGTCTGTGTGCTGGGATTCACATCTCTGCGCCCTTCAATGCTTCTC 3779  
QY 1259 rSerValLeuPheArgProIleHisProValGIYGIYLeuArgCysArgCysProProGI 1279  
DB 3780 TTCGGGTCTTCCGGCCCATCACTCGTCGAGAGGTCTGCGCTGTGCTGCCACCCGG 3839  
QY 1279 yPheThrGIYAspTrpCysGIYThrGIYValAlaPheLeuCysIYrSerArgProCysGIYPr 1299  
DB 3840 CTTAACAGGCGACTACTGTGAGACCGAGGTGACCTCTGTACTCAAGGACCTTGGAGAC 3899  
QY 1299 oHIsGIYArgCysArgSerArgGIYGIYGIYTrpThrCysLeuCysArgAspGIYTrpThr 1319  
DB 3900 CCATGGCGCTGCGGAGCGAGGAGGTGGCTATACCTGCTGTGCTGAGATGCTACAC 3959  
QY 1319 rGIYGIYHisCysGlnValSerAlaArgSerGIYArgCysThrProGIYValCysIYHis 1339  
DB 3960 GGGCGAGCACTGTGAAGCAGATACCACTCAGGCCCTGTACTCCAGGTGTGCGACAGA 4019  
QY 1339 nGIYGIYThrCysValAsnLeuLeuValGIYGIYPheIYHisCysAspCysProSerGIYHis 1359  
DB 4020 CGGGGTACTGTGTACACTGTGTGAGGAGGTATAAAGTGTGACTGCCATTCGGGGA 4079  
QY 1359 pPheGIYArgProIYArgGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
DB 4080 CTTTGAAGAAACCTTGTGCAAGGTGACACAGCACTTCCCTGCGGCCCTTATATCAC 4139  
QY 1379 rPheArgGIYLeuArgGlnArgPheHisPheThrIleuAlaLeuSerPheAlaThrIYHisGI 1399

DB 4140 CTTCCGTGGCTGACACAGCGCTTCACTTCACTGCGCTTCTGCTTCTTACCAAGAA 4199  
QY 1399 uArgAspGIYLeuLeuLeuTrpArgGlnArgPheAsnGluYHisAspPheValAlaIle 1419  
DB 4200 GCGTATACGGCTACTGTGTATCAATGGGCGCTTCAATGAGAAAGCACTTGTGGCTCT 4259  
QY 1419 uGluValIleGlnIleuGlnValGlnIleuThrPheSerAlaGIYGIYLeuSerThrThrVal 1439  
DB 4260 CGAGGTATTCAGAAACAGGTGAGCTCACCTTCTGCAAGGGAATTCAGACACACCGT 4319  
QY 1439 lSerProPheValProGIYGIYValSerAspGIYGIYTrpHisThrValGlnLeuIYHis 1459  
DB 4320 GTCTCCGTTTGTGCGGAGGAGGTCAAGTATGGCCATGGGCACTGCTGAGACT 4379  
QY 1459 rTyAsnIYsProLeuLeuGIYGIYTrpGIYLeuProGlnIleProSerGIYGIYHisVal 1479  
DB 4380 CTATATACAGCCCACTGTGGGTGACAGAGGCTTCCCAAGTCCATCTGAGCAGAAAGT 4439  
QY 1479 lAlaValIleThrValAspGIYCysAspThrGIYValAlaLeuArgPheGIYSerValIle 1499  
DB 4440 AGCGGTGTGCTGAGATGCTGTGACACAGGGGTGCTTGTGCTTGGAGCTATGCT 4499  
QY 1499 uGIYAsnTrpSerCysAlaIleGlnGIYTrpGlnGIYGIYSerIYsSerLeuAspLe 1519  
DB 4500 GGGCACTACTCTGTGCTGCTGCCAGGGCACTCCAGAGGAGCAGAGAAAGTCTTGCACCT 4559  
QY 1519 uThrGIYProLeuLeuLeuGIYGIYValProAspLeuProGluSerPheProValArgMe 1539  
DB 4560 GACAGGGCCCTGTGTGTGGGTGGGTGGCGGATCTCCCAAGAGCTTCCCGTCCCAAT 4619  
QY 1539 rArgGlnPheValGIYCysMetArgAsnLeuGlnValAspSerTrpHisIleIleAspMetAl 1559  
DB 4620 GCGGCACTTGTGGGCTGATGAGAGACCTCCAGGTGATAGCGGCACTGACATGAGGC 4679  
QY 1559 aAspPheIleAlaAsnAsnGIYThrValProGIYCysProAlaIleYHisAsnValCysAs 1579  
DB 4680 CGACTTACTTCCAAACATGGCACTGTGCTGCTGGCCCAACAAAGATCGTGTGCA 4739  
QY 1579 pSerAsnThrCysHisAsnGIYGIYThrCysValAsnGlnTrpAspAlaPheSerCysGI 1599  
DB 4740 CAGCAAGCTTGTGCCAACAGGTGGCACCTGTGTAAACAGAGAAATCGTTCAGCTGGCA 4799  
QY 1599 uCysProLeuGIYpPheGIYGIYIYsSerCysAlaGlnIleMetAlaAsnProGlnHisPhe 1619  
DB 4800 GTGTCCACTAGGCTTCCGGGGCAAGACTGTGCCCAAGAAATGGCCAAACCCCAAGCTTT 4859  
QY 1619 eLeuGIYSerSerLeuValAlaTrpHisGIYLeuSerLeuProIleSerGlnProTrpTy 1639  
DB 4860 TCTGGGCGAGCAGCTTGTGAGCTGAGTGGCTCTATCTACCAATCTTCAAGCCCTGGCA 4919  
QY 1639 rLeuSerLeuMetPheArgThrArgGlnAlaAspGIYValLeuLeuGlnAlaIleThrAr 1659  
DB 4920 CCTCACTCAATGTTCGACACACCGCAGGAGAAAGGTCTTACTGTGACGGCTGCACAG 4979  
QY 1659 gGIYArgSerThrIleThrLeuGlnIleuArgGlnGIYHisValIleMetLeuSerValGIYGI 1679  
DB 4980 GGGGCGAGCAACATACCTCGAGCTTCCGGGAGGCCAAGTGAAGGTCAATATAGAGAGG 5039  
QY 1679 yThrGIYLeuGlnAlaSerSerLeuArgLeuGlnProGIYArgAlaAsnAspGIYAspTr 1699  
DB 5040 CACAGGGCTCCAGGCGTCACTCTGTGCATCTTGAACCGAGCGCAGACCAATGATGTGACTG 5099  
QY 1699 pHisIleAlaGlnIleuAlaLeuGIYAlaSerGIYGIYProGIYHisIleAlaIleLeuSerPh 1719  
DB 5100 GCATATACAGAGCTTGGGACTGGAGACCGGAGGCTTGGCCATGACATTCGTCTT 5159  
QY 1719 eAspTrpGIYGIYGlnArgAlaGlnGIYAsnLeuGIYProArgLeuHisGIYLeuHisIle 1739  
DB 5160 CAACATATGGCAACAGACAGAGGATATCTGGGCTTGGCTGATGAGTGCACCT 5219  
QY 1739 uSerAsnIleThrValGIYGIYIleProGIYProAlaGIYGIYValAlaArgGIYPheAr 1759



Db 5220 GAGCAATATTACAGTCGGGGAGTTCTGGGCGACCGAGAGTGGCCCGCTTCG 5279  
Qy 1759 GGIYCySeIuengIuNgIyValArgvalSerAspThrProgluGIyValaIaenSerLeuAspPr 1779  
Db 5280 GGGCGTTTGACAGGGGTGTCAAGGTTAAGCAAGACCTTGAAGGGGTTCACAGTCTAGATCC 5339  
Qy 1779 oSerHieGIyGIuSerIleAenValGIuGIuNGIyCySeSerLeuProAspProCyAspSe 1799  
Db 5340 CAGCGCGGGAGAGCATCAATGTGGAGCCAGGCTGTAGCCCTGCAGACCCCTGTGACTC 5399  
Qy 1799 rAsnProCyProAlaenSerIyCySeSerAsnAspTrpAspSerIySerCySeSerCy 1819  
Db 5400 GAATTCAGTCCCTTACCAACAGCTACTACAGCAATACAGTGAAGACCTTCTGTAGCTG 5459  
Qy 1819 sAspProGIyTrpYrGIyAspAsnCyThrAsnValCyAspLeuAsnProCySeGIuH1 1839  
Db 5460 TGTTCAGTACTATAGTGTGACCACTGTACAAATGTGTGTGACCTGAGCCATGAGGACA 5519  
Qy 1839 sGIuSerValCyThrArgIySerProSerAlaProHieGIyTrpCySeGIuCySeProPr 1859  
Db 5520 CAGTCCGTGTACCCGAAAAACCAATACACCCACGCTTACATCTGCGAGTGTTCAC 5579  
Qy 1859 oAsnTrpLeuGIyProIyCySeGIuThrArgIleAspGIuProCySeProAspGIyTrpTr 1879  
Db 5580 AAATTACTTGGGCCATATTGTGAGACCAAGATTGACCAACTTGGCCCGGTGGTGTG 5639  
Qy 1879 pGIyHieProThrCySeGIyProCyAsnCyAspValSerIySeGIyPheAspProAspCy 1899  
Db 5640 GGGACACCCCAATGTGTGCATGCACTGTGATGTGCAAAAGGCTTGTGACCCAGATGTG 5699  
Qy 1899 sAsnTrpSerIyGIuCySeHieCySeGIuGIuAsnHieTrpYrTrpProProGIySeSerPr 1919  
Db 5700 CAACAGACCAAGTGGAGGTCCACCTGCAAGAGAAAGCATACCGGCCCGGACAGCC 5759  
Qy 1919 oThrCySeLeuLeuCyAspCySeIyTrpProThrGIySerLeuSerArgValCyAspAspProGI 1939  
Db 5760 CACTGTCTCTGTGTGACTGTACCCCACTGGTCTTGTGCCAGTGTGTGACCCAGCA 5819  
Qy 1939 uAspGIyGIuCySeProCySeIyAspProGIyValIleGIyArgIuCyAspAspCyAspAs 1959  
Db 5820 GGAAGGCGCAGTGTCCGTGCAAGCCGTGAGTGTGGGGCTGAGTGTGACCGCTGTGACAA 5879  
Qy 1959 nProPheAlaGIuValThrTrpAsnGIyCySeGIuValaAsnTrpSerSerCySeProAlaGI 1979  
Db 5880 CCGTTTGTGTAGGTACACCAACAAAGTGTGAAGTGAATTAACAACAGCTGCCACAGGCG 5939  
Qy 1979 aIleGIuAlaGIyIleTrpTrpProArgThrArgPheGIyLeuProAlaAlaIaIaProCy 1999  
Db 5940 CATAGAGGCTGGGATCTGTGGCCCGGACCGGCTTCCGGCTACTGTGCTGCCCTG 5999  
Qy 1999 sProIySeGIySerPheGIyThrAlaValArgHieCyAspGIuHieAspGIyTrpLeuPr 2019  
Db 6000 CCCCCAAGGCTCCCTTGTGGACTGTGTGGCCCACTGTGATGAGACAGGGGTGGCTGCC 6059  
Qy 2019 oProAsnLeuPheAsnCyThrSerIleThrPheSerGIuLeuGIySeGIyPheAlaGIuAr 2039  
Db 6060 CCGAACCTCTTCACTGACGTGACGTACCTTCTCAAGAACTAAAGGGCTTGTGTGAGCG 6119  
Qy 2039 gLeuGIuArgAsnGIuSerGIyLeuAspSerGIyArgSerGIuNGIuLeuAlaLeuLe 2059  
Db 6120 GCTGCAGAGGACCAATCGGGCTGTGACTCAAGACGCTCCCAAGAGGTTAGCCCTGTCTCT 6179  
Qy 2059 uArgAsnAlaThrGIuHieThrAlaGIyTrpPheGIySerAspValIySeAlaIaIaIyGI 2079  
Db 6180 GCGCAATGCCACCAAGACCACTGTGGCTACTTGGCAGTGTCAAGGGTGGGCTTACCA 6239  
Qy 2079 nLeuAlaThrArgLeuLeuAlaHieGIuSerThrGIuArgGIyPheGIyLeuSerAlaIaTh 2099  
Db 6240 GCTGGCCACAGGGCTTGTGCTCAAGAGTCCCAAGGGGGCTTGTGGCTATCCGCGCAC 6299  
Qy 2099 rGIuAspValHiePheThrGIuAsnLeuLeuArgValGIySeAlaLeuLeuAspThrAl 2119  
Db 6300 ACAAGAGTACACTTACCGAGGAATCTGTGAGGGTGGGAGGCCCTCTCTCGATGTGAGC 6359

Qy 2119 aAsnTrpArgHieTrpGIuLeuIleGIuNGIuThrGIuGIyGIyThrAlaTrpLeuLeuGI 2139  
Db 6360 CAACAGAGGCACTGGGAACTGATTCAGACAGAGAGGGTGGACCCGCTGTGGCTTCCA 6419  
Qy 2139 nHieTrpGIuAlaTrpAlaSerAlaLeuAlaGIuAsnMetArgHieThrTrpLeuSerPr 2159  
Db 6420 GCACTATGAGGCTTACCCAGCGCCCTTGGCCAGAAATATGGGCAACCTTACTTAAGGCC 6479  
Qy 2159 oPheThrIleValThrProAsnIleValIleSerValArgLeuAspIyAspGIyAsnPh 2179  
Db 6480 CTTACCACTGTACACCCCACTTGTCAATCTGTATGTGAGCTGGAGTAAAGGGAACTT 6539  
Qy 2179 eAlaGIyAlaIyLeuProArgTrpGIuAlaLeuArgGIyGIuNGIuProProAspLeuGI 2199  
Db 6540 TGTGGAGCAAGCTGCCCGTTATGAAAGCATTAAGTGGGAGAGGCCCCCAAGATGTGA 6599  
Qy 2199 uThrThrValIleLeuProGIuSerValPheArgGIuThrProProValAlaTrpAla 2219  
Db 6600 GACCAAGTCAATTTGTGCAGAAATCTGTGTCAAGAGATGCCCCCTCCATGTGTAGTGTGC 6659  
Qy 2219 sGIyProGIyGIuAlaGIuNGIuProGIuGIuLeuAlaArgArgGIuArgArgHieAspProGI 2239  
Db 6660 AGAAGCTGTGAAAGCCAGAGAGAGAGAGCTGGCCCGGCGAGCGAGGACCCAGCA 6719  
Qy 2239 uLeuSerGIuGIyGIuAlaValAlaIaSerValIleIleTrpArgThrLeuAlaGIyLeu 2259  
Db 6720 ACTGAGTCAGAGAGGCACTGGCCAGTGTCAATCATCTACCAACCTGTGGTGACTGCT 6779  
Qy 2259 uProHieAsnTrpAspProAspIyAspSerLeuArgValProIyAspProIleIleAs 2279  
Db 6780 GCCCAACAACTATGACCCCGATAGCGGTAGCTTGAAGTGTCCCAAGGCCCGCTCATCAA 6839  
Qy 2279 nThrProValaIleSerIleSerValHieAspAspGIuNGIuLeuLeuProArgAlaIaLeuAs 2299  
Db 6840 CACCCCGCGGTAGATCAAGCTTCAACAGACAGAGAACTCTGCCGGGGCTGTGGA 6899  
Qy 2299 pIyAspProValThrValGIuPheArgLeuLeuGIuThrGIuNGIuArgTrpIleCy 2319  
Db 6900 CAAGCCAGTACCGGTGAGTTCGGGCTGTGAGAGCCAGAGAGGAGCAAGCCATCTG 6959  
Qy 2319 sValPheTrpAsnHieSerIleLeuValSerGIyThrNGIyGIyTrpSerAlaArgIyCy 2339  
Db 6960 TGTCTTCTGGAACATTCATCTGTGTCACTGTGAGTGGCTGGGCTGCCCGAGGCTG 7019  
Qy 2339 sGIuValaIaPheArgAsnGIuSerHieValSerCySeGIuAsnHieMetThrSerPh 2359  
Db 7020 CGAAGTGTCTTCCGTACAGAGCCACGTCACTGCCAGTGCACCAATATGACAGCTT 7079  
Qy 2359 eAlaValaIeMetAspValSerArgArgGIuAsnGIyGIuIleLeuProLeuIyThrIle 2379  
Db 7080 TGCATGTCTTATGATGATGTCAAGAGAGAGAGCGGTGAGATCTTGTGCCATGAAGACCT 7139  
Qy 2379 uThrTrpValAlaLeuGIyValThrLeuAlaIaLeuLeuLeuThrPhePheLeuTh 2399  
Db 7140 GACATAGTGTGGCCCTTGTGAGTCACTTGTGCTGCCAGTGTCACTCTCTTCTCTCAC 7199  
Qy 2399 rLeuLeuArgIleLeuArgSerAsnGIuHieGIyIleArgArgAsnLeuThrAlaIaLe 2419  
Db 7200 CTTCTCTCGGCGCCCTCCGTCCCAACAGACAGGCACTCCAGCACTTCACTGCTGCT 7259  
Qy 2419 uGIyLeuAlaGIuLeuValaPheLeuLeuGIyIleAsnGIuAlaAspLeuProPheAlaCy 2439  
Db 7260 GGGGCTGGCCCAAGCTGTCTTCTCTGGGCAATCAACAGGCTGAGACTCTCTTGTGCTG 7219  
Qy 2439 sThrValIleAlaIleLeuLeuHiePheLeuTrpLeuCySeThrPheSerTrpAlaLeu 2459  
Db 7320 TACAGTATGTGCTATCTGTGTGATTTCTGTACCTTGTCACTTCTCTGGGCTGTGCT 7179  
Qy 2459 uGIuAlaLeuHieLeuTrpArgAlaLeuThrGIuValaArgAspValaAsnThrGIyProme 2479  
Db 7380 GGAAGGCTTACACTTGTACCGGCGCTCAAGAGGTGGGAGTGTCAATGCCAGTCCAT 7439

2479 CARGPHERYTYRMETLEUGLYTPGLYVALPROLALPHEHLETHRGILYLAVALGL 2499  
 7440 GCGTTTCTACTACAGCTGCGGCGTCCCTCTTCATCAAGGCTCGCTGCGG 7459  
 2499 YLEUAPPROGUGLYTYRGLYANPRAAPHECYETRPLEUSERILEYTPAAPTHTLE 2519  
 7500 CTGGATCCGAGAGGCTATGGAGACCCTGACCTTCGCTGCGCTCTGTGCTATGAACTT 7559  
 2519 UILETRPSEAPHEALAGLYPROVALALAPHEALVALSERMETSERVALPHELEUTYL 2539  
 7560 CATCGAGGCTTGTGCGAGCAAGTGGCTTGTCTTTCCAGAGAGGCTTCTCTATCAT 7619  
 2539 ELEUHALAALAAAGALASERCYVALAALAGLARGLNGLYPHEGLULYLYVGLYPROVA 2559  
 7620 CCGTGGCGCCGAGGCTCTGTGCTGCGCCCAAGCGAGGCTTCGAGAAAGAGGCTCGT 7679  
 2559 ILSERGLYLEUGLNPROSEAPHEALVALLEUUEUUEUSERALATHTTRPLEUENAL 2579  
 7680 CTGGAGCTGCGCTCTCTTCACGAGTCTCTGCTGCTGAGTGCACAGTGCCTGCTGCGC 7739  
 2579 ALEUUEUSERVALAANSEAPRTHRLLEUUEPHEHLESTYRLEUAPHEALATHRCYASAPCY 2599  
 7740 ACTGCTCTCTGTCACAGTGAACACTGCTCTTCCACTACCTTCGCTGCTGCAATTTG 7799  
 2599 AILEGLNGLYPROPHLEUHEPHELEUSERTYRVALYALLEUSERLYVGLYVALYGLYAL 2619  
 7800 TGTCCAGAGGCTCCCTCATCTCTCTCAACGCTGTGCTGCAAGCAAGAGTCCGGAAGC 7859  
 2619 ALEUYSLEUALACYSEAPRGLYVPSERPROAPRPROALALAUTHTHTHLYSETH 2639  
 7860 ACTCAAGTTCGCTCTGACCGGAGGCCAGTCCGACCTGCTGCTGACCACTAATGATAC 7919  
 2639 ILEUHTHSESESTYRASNCPYPSERPROTYRVALAAPGLYARGLEUTYGLNPROLY 2659  
 7920 CCTGACCTCTGCTCTAATACGTGCCCCAGCCCTAAGCAAGCGAGGCTGATCAACCTTA 7979  
 2659 RGLYAPSERALAGLYSERLEUHEHLESETHSERARGSERGLYVPSERGLNPROSESTY 2679  
 7980 CGGAATTCGCTGCTGCTCTCAACAGTGCACGCGCTCCGCAAGAGTCAACCCAGCTA 8039  
 2679 IILEPROPHLEUUEUARGLUGLUSERALALEUENPPOGLYGLNGLYPROPGLYLE 2699  
 8040 CATCCCTCTTCTGCTGAGGAGAGTCCAGCTGAACCTGCGCCAGTTCCTCCCTGCGCT 8099  
 2699 UGLYAPPROGLYSERLEUHELEUGLUGLYLAPAGLNGLNIASAPPROAPRTHAS 2719  
 8100 AGGGAGCCCAAGCGGCTGTCTCTGGAAGGCCAACCCCAACAATGATCCGACACGGA 8159  
 2719 NSEAPSERAPLEUSERLEUGLUBAPRGLINSEGLYSESTYRVALASERTHTHISE 2739  
 8160 CTCTACAGTGAACCTGCTCTGGAAGATGACAGAGTGTCTCTATGCTTCCACCTCATCTC 8219  
 2739 ISEAPSERGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 2759  
 8220 ATCAGACGAC-----GAGGAGAGAGAGAGAGAGGCGCTTCCCTGCGGAGCA 8267  
 2759 NGLYTPAPSERLEUGLYPROGLYVALAGLULARGLEUPROLEUHSERTHTPROLY 2779  
 8268 GGGCTGCGGACAGCTGCTGCTGCTGAGAGTGAAGACTGCCCCCTTCAAGTACCCCA 8327  
 2779 SAPRGLYGLYPROGLYPROGLYLYSALAAPROTHPRPROGLYAPRPHGLYTHTHALY 2799  
 8328 GATGAGGAGTCCGGGCTCTGGAGAGGTCCCTTGGCTGGAGACTTGGAGCAACAACAA 8387  
 2799 SGLUSERSERGLYAMGLYVALAPROGLUGLULARGLEUARGLUGLUGLUGLUGLUGL 2819  
 8388 GAGAGATAGTGTAGTGGGCTTGAAGAGGCGCAAGGAGGAGTGAAGTGGCTTAAC 8447  
 2819 TARGLUGLYSERLEUGLYPROLEUPROGLYSESERALAGLNPROHIALYVGLYILE 2839  
 8448 TCGGAGAGGCTCTGAGGACCCCTTCCGGGCTCTTCAACCACTCAACAAAGGATCTCT 8507  
 2839 ULYSYLSYSCYBLEUPROTHILSESTGLYYSERSESTLEUENALARGLEUPROLEUGL 2859

Db 8508 CAAGAGAGTGTCTGCCACCATCACAGGAAAGAGGCTCTTAAGGCTGCCCTCGGA 8567  
 QY 2859 UGLNCSETHRGYSESERARGLYSERSERALASERGLUGLYSERARGLYPROPR 2879  
 Db 8568 GCAAGGACAGAGGCTCTTCTGCGGCTCTCTTCATCAAGAGGCGAGTGGCAGCACTCC 8627  
 QY 2879 OPRARGPROPROARGLINSERLEUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 2899  
 Db 8628 TCCCGCCCACTCGCGCCGCTCTGCAAGAAACACTGAATGGGCTCATGCTGCTGCGC 8687  
 QY 2899 AMESERTILEYALAGLYTHRVALAAGLUBAPRSESTYRGLYSESTGLUPHELEUPHEP 2919  
 Db 8688 CATAGAGATCAACGACGACAGCGAGTGAAGTGAAGTCTCTGCTGCTCAATTTCTTCTT 8747  
 QY 2919 EASPHLEUHS 2923  
 Db 8748 TAACTTCTGCTCAT 8760

## RESULT 7

D87469 8924 bp mRNA linear PRI 06-OCT-2001  
 LOCUS Human mRNA for KIAA0279 gene, partial cds.  
 DEFINITION  
 ACCESSION D87469  
 VERSION D87469.1 GI:1665820  
 KEYWORDS KIAA0279.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

## REFERENCE

1 Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y., Ohara, O., Tanaka, A., Kotani, H., Miyajima, N. and Nomura, N.  
 Prediction of the coding sequences of unidentified human genes. VI.  
 The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 analysis of cDNA clones from cell line KG-1 and brain

JOURNAL DNA Res. 3 (5), 321-329 (1996)  
 MEDLINE 97191544  
 PUBMED 9039502

## REFERENCE

2 (bases 1 to 8924)  
 AUTHORS Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (27-AUG-1996) Osamu Ohara, Kazuwa DNA Research Institute;  
 1532-3, Yama, Kisarazu, Chiba 252-0612, Japan  
 (E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913)

## FEATURES

location/Qualifiers

source

1..8924  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HA6133"  
 /sex="male"  
 /tissue\_type="brain"  
 /clone\_id="pSPORT 1"  
 1..8924  
 /gene="KIAA0279"  
 <1..7227  
 /gene="KIAA0279"  
 /note="Similar to D.melanogaster cadherin-related tumor suppressor"  
 /codon\_start=1  
 /protein\_id="BA13407.1"  
 /db\_xref="GI:1665821"

## CDS

/translation="ESVPLGLVLAHQAIDADAGDNRLRYRLAGVGHDPPTINNGT  
 GWTSLAELREERDPSFVBARDHGTPTALTAASVTVLVLDVNDNPPFTQPEYTV  
 RINEDAVGTSVTVSAVDRAHSAVITYQTSGTNRPFSTSGGGLVSLAPLDY  
 KLEQYVAVTASGTRQDRAQIVVNTDANTHRPVQSHYTVNVNEDRPAGTVVL  
 ISATDPTGNARITPMEDSIQFRIDADGAVTTOAARDIYEDQVSTTLAATARDG  
 IPQSDPTTYELIYVNDNAPQFLRDSYGSYEDVPPTTSVTQISATTRDSATNR  
 VFTFGGDDGDFIVESTSGIVTKLRIDRENAQYVLAIVADGMPPTKMEV  
 TVVTVLDVNDPVPVEODEPVEENSPVIGLAVARVATPDDEGTNAQIMQIVEGNI  
 PEVQDLDFSGELTALVDLYBDRPEYVLVQATSAVLVSRATVHVALBRNDPPTV  
 GNFEILNNVTNRSSSPGGAIGRVAPHPDISDSILTSFERGNEISLVTLNASTGE

LKLSBALNNRPLEAIIKSVLSDGVHVTQAQALRTIIIDEMLTHTSITTLRLBDMSPER  
 RFLSEDLPIQAVATLATTDPDHYVNVQVODTDAPGHIILNTSLVSGQPPGGGPP  
 PFLSEDLQERLYLNRSLITLTAISAQVLPDPDNICLRPCENKRYCVAFLPSSAPF  
 IASSSVLPFRPHVGLRCRCPPGTGYCETBYDLGYSRPGPHGRCSREGGVTTCL  
 CRDGYTGHCEVASASRGCTPGVCKNGCTCNLLVGRKCCPSGDPPEKPYCTTTS  
 PPASPRITFRGLRORPHITLALSPATKRDGILLYNGRPNRKHDPVLEVOYVOLT  
 FSAGSITTVSPFVPGVSDQGMHTVQIKYNNKPLLGQTVLPQSGEBOKAVVTVQD  
 DTGVALRPGSVLGANSYCAQGTQGGSKSLDTGFLLAGVPLDPSPFPVAMRQPVGC  
 MRNLQVDSRHIDMDPIANNGTGPGCPAKKVCDSNTCHNGGTGVNOMDASCEPLG  
 FGKSCAQEMANPOHFLGSLVAMHGLSLPISOPMYLIMRTROADGVLLQATYRGR  
 STITOLAREHVMLSVEGTGLQASSLRLRBRANDGMHHAOLGASGGGHNLSP  
 DYGOORAGNGRPLRHGHLNITVGGIPRPGVYARPRGCLQGVTSIDTBNVCLNL  
 DPSSGESINVEQCSLPDPCSNPCPNASVCSNMDYS CSCDPGYEDNCTNCDNL  
 PCHEQSVCTKRSPAHGYTCRPPNYLAPYCETRLIDQCPGMMGHPCGCGCNDSVK  
 GFDPCKNTSGECHKENHRYRPGSPYTLCDQPYTGLSYVCPEDQCCPKGVIG  
 ROCBRCNDPFAVETNGCEVNVYDSCPRALBGMWRPRFGIPAAAPFKGSFGTAVR  
 HCDERHGMPLNLFNCTSTITPSLTKGPRALORNSGQISDGRSOOLALLNANQHTA  
 GYFGSDVYVAVQOLATRLAHSTQGRFSLATQVHTFENLRVGSALLDTRAKHME  
 LIQOTBEGTAMVLOHTEAVASALAQNMHRTYLSPTVTPNIVISVARDGNRAGAK  
 LPRYALRGEQPPDLETTVILPESEYRETPPVVRPAGBEOBELARQRHPELS  
 OGEAVASVITRTLAGLPHNYDPDKRLAPKRPINTPVVISVHDEELRLPDL  
 KPTVQFPLETBERTEKPCVEMNHSILVSGGWSARGCEVFRNESHVSOCNHT  
 SPAYMDVSRRENGELPLKTLTYVALGVTLLAALLPPELTLLRIEENOHGRYRL  
 TLAAGLADVPLIGINQNDLPRACTVIALIHLFYLCTPSNALLBALHTALTEVRD  
 VNTGPMRTYMLGMPVPAFITGLAVLDPBEGVNDPFWLSIYDTLMSFAPVAVAY  
 FMSVFLYILAAASCAQORQEKKPGVSGLOPSPAVLLLSATMLLSVNSDTLL  
 SHYFATCNCIQGPPIFLSYVLSKEVKALKACSRSPDPAITLSTLSTNSCP  
 SPYADGRLYOPRGDAGSLHSTSRGKQPSYIPILAREBALNGQOPGAGDMSL  
 FLEGGDOHDHDDSDSLSLBDOSSGYASTHSDSREEREEREAAPGEGCGMSL  
 LLGRABERLPHLSTPKDQGRPGKAPMGEDGTTRAKESGNGAPREERLRENGDLNRS  
 GSLEGPRESSAQPHKGLIKKCLPTISKSSILRLPRLPESLSSGSSASSGSGGPP  
 PRPPERQSLQLOQNMPIAMSIKAGTYEDBDSGSEPLFNFHLS<sup>1</sup>

BASE COUNT 1727 a 2846 c 2504 g 1847 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 8924  
 Score: 12833.00 Matches: 2408  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 82.55% Indels: 0  
 DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x D87469 (1-8924)

QY 516 GluservValProleuGlyTyrleuValleuHisValGlnAlaIleAspAlaAspAlaGly 535  
 DB 1 GAGAGCGTCCCTTAGGCTACCTGTTCTCCATGTCGAGGCTATGACGCTGATCTGCT 60  
 QY 536 AspaenAlaArgleuGluTyrArgleuAlaGlyValGlyHisAspPheProPheThrIle 555  
 DB 61 GACATATGCCCGCTGGAATACCGCCTTGCTGGGGTGGGACATGACTTCCCTTCCACCATC 120  
 QY 556 AsnaenGlyThrGlyTyrTrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPhe 575  
 DB 121 AACCATGGCAACGAGCTGATCTCTGTGGCTGCTGAACCTGGACCGGAGAGAAAGTTGATTTTC 180  
 QY 576 TyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSer 595  
 DB 181 TACAGCTTTGGGGTGAAGAGCTCGAAGCATGGCACTTCCAGCACTGACTGCTCGGCGAGT 240  
 QY 596 ValSerValThrValLeuAspValAsnAspAsnaenProThrPheThrGlnProGluTyr 615  
 DB 241 GTCAACGTGACGTCTGCTGATGTCAACGACACATTCACCTTAAACCAACCAAGATAC 300  
 QY 616 ThrValArgleuAsnGluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaAla 635  
 DB 301 ACAAGTGGCGCTCAAGTGAAGATGACGTGTGGCACCAAGCGTGTACCGTGCACCTGTG 360  
 QY 636 AspaTgaAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArg 655  
 DB 361 GACCTGATGCTCATAGTGTCTATCACCAGATCACCAAGTGGCAATRACCTGGAACCGC 420

QY 656 PheSerIleThrSerGlnSerGlyGlyGlyLeuValSerleuAlaLeuProleuAspTyr 675  
 DB TTTCTTCATACCAACCAAGTGGTGGTGGTATTCCTTGGCCCTGACCTGAGATAC 480  
 QY 676 LysleuGluArgGlnTyrValleuAlaValThrAlaSerAspGlyThrArgGlnAspThr 695  
 DB 481 AAATCTTGAAGCGCGAGTATGTTGGCTGTGATTAACCCCTCCGATGACCTCGGACAGACAG 540  
 QY 696 AlaGlnIleValAlaAsnValThrAspAlaAsnThrHisAspProValPheGlnSerSer 715  
 DB 541 GCACAGATTGGTGGAATGTCACCGACCGCACCAACCATGCTCTGCTTTCACAGCTCC 600  
 QY 716 HisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValleuIleSer 735  
 DB 601 CACTATACAGTGAATGTTAATGAGAACCGGCGGACAGCACACAGGTGTCTGATCAGC 660  
 QY 736 AlaThrAspGluAspThrGlyValAsnAlaArgIleThrTyrPheMetGluAspSerIle 755  
 DB 661 GCCACGAGATGAGACACAGAGTGAATGCCCGCATACCTTCAATGAGAGACAGATC 720  
 QY 756 ProGlnPheArgIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyr 775  
 DB 721 CCCCATGTTCCGATCGATCGACAGACCGGGGCTGTACACCCAGGCTGAGCTGACTAC 780  
 QY 776 GluAspGlnValSerTyrThrleuAlaIleThrAlaArgAspAsnGlyIleProGlnLys 795  
 DB 781 GAAGACCAAGGTCTTACACCTGCGCATTAAGTCTCTGGGACATATGCAATGCCATCCCAAG 840  
 QY 796 SerAspThrThrTyrleuGluIleleuValAsnAspValAsnAspAlaProGlnPhe 815  
 DB 841 TCCCAACCAACCTACCTGAGATCTCTGTGAACCACTGAATGACATATGCCCTTCAGTTC 900  
 QY 816 LeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerValleu 835  
 DB 901 CTGCGAATCTCTACCAAGGCGCATGTCTATGAGATGTGCCACCTTCACTAGGTCTGTG 960  
 QY 836 GlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGln 855  
 DB 961 CAGATCTCAGCACTGATCGATTCGTGATTCGATTAATGGACGGGTCTTCAACCTTCAA 1020  
 QY 856 GlyGlyLeuAspAspGlyAspAspPheIleValGluSerThrSerGlyIleValArgThr 875  
 DB 1021 GGAGGCGACATGAGGAGCGGTGACTTATGTTGATGTCACGTGACGATGTGCGAAGC 1080  
 QY 876 LeuArgArgleuAspArgGluAsnValAlaGlnTyrValleuArgAlaTyrAlaValAsp 895  
 DB 1081 CTACGAGAGCTGATTCAGAGAGAGCTGCGCCACGATGTCTTGGCGGACATAGAGCTGAC 1140  
 QY 896 LysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAspValAsn 915  
 DB 1141 AAGGGGATGCCCGCCAGCCCGCACCTATGGAAGTGAACGTCACTGTGTGGAATGTAAT 1200  
 QY 916 AspaenProProValPheGluGlnAspGluPheAspValPheValGluLysSerPro 935  
 DB 1201 GACATATCCCTGCTTGTAGCAGAGATGATGATGTTGTGGAGAAAGAACAGCCCC 1260  
 QY 936 IleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGln 955  
 DB 1261 ATTGGGCTAGCGCGTGGCGCGGGGTCAAGCCACCTGACCCCATGAAGCAACAGGCCAG 1320  
 QY 956 IleMetTyrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSer 975  
 DB 1321 ATTATGATCCAGATTTGAGAGGGGCAATATCCCTGAGGTCTTCCAGCTGAGACATCTTCC 1380  
 QY 976 GlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyrValleuVal 995  
 DB 1381 GGGAGGCTGACAGCCCTGAGTGAAGTGAAGTGAAGACAGGCGCTGAGTGAAGTCTGCTC 1440  
 QY 996 IleGlnAlaThrSerAlaProleuValSerArgAlaThrValHisValArgleuLeuAsp 1015  
 DB 1441 ATCCAGGCGACATGATGCTCTCTGTGAGCGGGCTACAGTCCAGTCCGCTCTTGAC 1500  
 QY 1016 ArgAsnAspAsnProProValleuGlyAsnPheGluIleleuPheAsnAsnTyrValThr 1035

Db 1501 CGCAAGACACACACAGTGGGCACTTGAATCCTTTCAACAACATATGCAAC 1560  
 Qy 1036 AenAArgSerSerPheProGlyValAlaIleGlyArgValProAlaHisAapProAap 1055  
 Db 1561 AATCGCTCAAGACGCTTCCCTGGGGGGTGCATTGGCCGAGTACTCTGCCATGACCTTGAT 1620  
 Qy 1056 ILeSerAapSerLeuThrTySerPheGluArgGlyAsnGluLeuSerLeuValIleLeu 1075  
 Db 1621 ATCTCAGATAGTGTGACTTACAGCTTTGAGCGGGGAAATGATCACTCAGCCGTGCTCGCTC 1680  
 Qy 1076 AenAlaSerThrGlyGluLeuGlyLeuSerArgAlaLeuAapAenAapProLeuGlu 1095  
 Db 1681 AAGGCTCCACCGGTAGCTGAAGCTAAGCCGCGCACTGGACAACACCGGCTCTGGAG 1740  
 Qy 1096 AlaIleMetSerValIleuValSerAapGlyValHisSerValThrAlaGlnCyValAla 1115  
 Db 1741 GCCATATGACGCTGCTGATGTAAGACGGGTACACAGCGTACCGCCGACGTGCGGCTG 1800  
 Qy 1116 ArgValThrIleThrAapGluMetLeuThrHisSerIleThrLeuArgLeuGluAap 1135  
 Db 1801 CGGTGACCATCATCAACGATGATGATGCTCAACCAAGCATCAAGCTGGGCTGGAGGAG 1860  
 Qy 1136 MetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAla 1155  
 Db 1861 ATGTCAACCGAGGCTTCTGTCAACACTGCTAGGCTCTTCATCCACAGGCGGTGGCGCC 1920  
 Qy 1156 ThrLeuAlaThrProAapPheHisValValIleAapValGlnArgAapThrAapAla 1175  
 Db 1921 AGCTGGCCACGCCACCGACACAGTGGTGGTCTTCAACCTACAGGGAGAACACCGAGCC 1980  
 Qy 1176 ProGlyGlyHisIleLeuAenValSerLeuSerValGlyLeuProProGlyProGlyGly 1195  
 Db 1981 CCGGGGGGCGACATCCTCAACGTAGCGTGTGGTGGGCGACGCCGCGAGGGCCGGGGG 2040  
 Qy 1196 GlyProProPheLeuProSerGluAapLeuGlnGluArgLeuTyLeuAenArgSerLeu 1215  
 Db 2041 GGGCGCCCTCTCTGCTCTGAGGACCTGCAGAGACGCTTATACCTCAACCGCACCTG 2100  
 Qy 1216 LeuThrAlaIleSerAlaGlnArgValIleuProPheAapAapAenIleCyLeuArgGlu 1235  
 Db 2101 CTGACGGCCATCTCGGACACGCCGCTGCTGCTTGGACGACAACTTCTGCTGGGGAG 2160  
 Qy 1236 ProCyGluAenTyMetArgCyValSerValIleuArgPheAapSerAlaProPhe 1255  
 Db 2161 CCTTGGAGAACTACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220  
 Qy 1256 IleAlaSerSerSerValIleuPheArgProIleHisProValGlyGlyLeuArgCyArg 1275  
 Db 2221 ATGCTCTCTCTCTCGGTGCTCTTCCGACCATCACCCGTCGAGGGCTGCGCTGCC 2280  
 Qy 1276 CyPProProGlyPheThrGlyAapTyArgGluThrGluValAapLeuCyTySerArg 1295  
 Db 2281 TGGCCGCCCGGCTTCAAGGTGACTATCGAGAACCGAGGTGACCTTCTGCTGCTGCTGCTG 2340  
 Qy 1296 ProCyGlyProHisGlyArgCyArgSerArgGluGlyGlyTyThrCyLeuCyArg 1315  
 Db 2341 CCTTGGGCGCCCAAGGCGCTGCCGACGCCGAGGGCGGCTACACTGCTGCTGCTGCTGCT 2400  
 Qy 1316 AapGlyTyThrThrGlyGluHisCyGluValSerAlaArgSerGlyArgCyThrProGly 1335  
 Db 2401 GATGGCTACACGGGTAGCACTGTAGGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2460  
 Qy 1336 ValCyLeuAenGlyGlyTyArgValAenLeuLeuValGlyGlyPheTyCyAapCy 1355  
 Db 2461 GTCTGACAGATGGGGGCACTGTGTCAACTGCTGTGGGGGCTTCAAGTGCATTGC 2520  
 Qy 1356 ProSerGlyAapPheGluArgProTyArgGlnValThrThrArgSerPheProAlaHis 1375  
 Db 2521 CCATCTGGAGACTTTCAGAGCCCTACTGCAAGTACCAACGCGAGCTTCCCGGCCAC 2580  
 Qy 1376 SerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrIleuAlaLeuSerPhe 1395

Db 2581 TCCCTCATCACTTTGCGGCGCTGCGCACGCGTTTCACTTCAACCTGGCCCTCTGCTT 2640  
 Qy 1396 AlaThrTyGluArgAapGlyLeuLeuLeuTyArgGlnArgPheAenGluTyHisAap 1415  
 Db 2641 GCCAAGAGAGACCGGACGGGTGCTGTGTTCATATGGCCGTTTCAATGAGAACATGAC 2700  
 Qy 1416 PheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLeuSer 1435  
 Db 2701 TTTGTGGCCCTGCAAGTATCCAGAGACAGGCCAGCTCACTTCTCTGCAAGGGAGTCA 2760  
 Qy 1436 ThrThrValSerProPheValProGlyValIleSerAapGlyGlnThrPheThrVal 1455  
 Db 2761 ACCACCAAGGTGTCCCATTCGTGCCCAGAGAGTCAAGTATGGCCAGTGGCATCGGTG 2820  
 Qy 1456 GlnLeuTyTyTyArgAenTyProLeuLeuGlyGlnThrGlyLeuProGlnTyProSer 1475  
 Db 2821 CAGCTGAATACTACATTAAGCCATGTTGGGTCAAGAGGCTCCACAGGGCCCATCA 2880  
 Qy 1476 GlnGlnTyValAlaValValThrValAapGlyCyAapThrGlyValAlaLeuArgPhe 1495  
 Db 2881 GAGCAGAAAGTGGCTGTGTGACCGTGAATGGCTGTGACACAGAGTGGCTTGGCTTC 2940  
 Qy 1496 GlySerValLeuGlyAenTySerCyAlaAlaGlnGlyTyThrGlnGlyGlySerTyHis 1515  
 Db 2941 GGATCTGTCTGGGCACTACTCTGTGCTGCCAGGGCACCCAGGTGGCAGCAAGAG 3000  
 Qy 1516 SerLeuAapLeuThrGlyProLeuLeuGlyGlyValProAapLeuProGlyLeuSerPhe 1535  
 Db 3001 TCTTGATCTGACGGGGCCCTGCTTACTTACGGCGGGGTGCTGACTGCTCCGAGAGCTTC 3060  
 Qy 1536 ProValArgMetArgGlnPheValGlyCyMetArgAenLeuGlnValAapSerArgHis 1555  
 Db 3061 CCAGTCCCAATGGGCACTTGTGGGTGACATGCGAATCTGCAAGTGAACAGCGCGCAC 3120  
 Qy 1556 IleAapMetAlaAapPheIleAlaAenAenGlyThrValProGlyCyProAlaTyHis 1575  
 Db 3121 ATAGACATGGCTGACTTATGTCACAAATGACACCGTGGCTGCTGCTGCTGCTGCTGCTG 3180  
 Qy 1576 AenValCyAapSerAenThrCyHisAenGlyGlyTyThrCyValAenGlnThrAapAla 1595  
 Db 3181 AACGTGTGACAGCAACTTGTCCACATGGGGGCACTTGGTGAACAAGTGGGACCG 3240  
 Qy 1596 PheSerCyGluCyPProLeuGlyPheGlyGlyTySerCyAlaGlnGluMetAlaAen 1615  
 Db 3241 TTCAGCTCGAGTGGCCCTGCTGGGCTTGGGGCAAGACTGCGCCCAAGAAATGCCCAT 3300  
 Qy 1616 ProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSer 1635  
 Db 3301 CCACGACCTTCTGGGACAGACCTGTGTGGCTGGCAATGGCTTCTGCTGCCATCTTC 3360  
 Qy 1636 GlnProTyTyLeuSerLeuMetPheArgThrArgGlnAapGlyValIleuLeuGln 1655  
 Db 3361 CAACCTGTGACTCAGCTCATATGTTCCGACAGGCGCAGGCGAGGTGCTGCTGTCAG 3420  
 Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeu 1675  
 Db 3421 GCCATCACACAGGGGGCGGACGACCTACCTACAGCTACAGAGGGCGCAAGTGAATGTG 3480  
 Qy 1676 SerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAen 1695  
 Db 3481 AGGTGAGAGGACAGGGCTTCAAGGCTCTCTCTCTGCTGAGACAGGCGCGGCAAT 3540  
 Qy 1696 AapGlyAapThrPheHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAla 1715  
 Db 3541 GACGGTGACTGGCACATGACAGCTGCACTGGGAGCGCAAGGGGGGCTGGCCATGCC 3600  
 Qy 1716 IleLeuSerPheAapTyGlyGlnGlnArgAlaGluGlyAenLeuGlyProArgLeuHis 1735  
 Db 3601 ATTCTGCTCTTGATTTAGGAGAGAGAGAGAGGAGGCAACTGGGGCCCCGGCTGAT 3660  
 Qy 1736 GlyLeuHisLeuSerAenIleThrValGlyGlyTyIleProGlyProAlaGlyGlyValAla 1755  
 Db 3661 GGTCTGACCTGAGCACTAACAAGTGGCGGAATCTTGGGCGACCGGGGGTGTGGCC 3720

OY	1756	ArgglYpHeaRgglCYeLeuGlInglYValArgValSerAapThrProGlInglYValAsn	1775
Db	3721	CGTGCGCTTTCGGGGCGCTGTTTGGAGGGTGTGGCGGCTGAACCGATACCGCAGAGGGGGTTTAC	3780
OY	1776	SerLeuAaPProSerHISglYgluSerLIleAenValISglInglYCySeSerLeuProAaP	1795
Db	3781	AGCTTGATCCAGCCATGAGGGAGAGGACATCAACGAGGACAAAGGCTGAGCCTCGCCGAC	3840
OY	1796	ProCYaAPSerAaPProCYaAProAlaAaSerTYrCYaSerAaAaPTrAaPSeTYr	1815
Db	3841	CGTTGTGACTCAAAACCGGATCTGTCTAAACGCTATTTCAGAAACGACTGGGACAGCTAT	3900
OY	1816	SerCYaSerCYaAPProGlYTYrTYrTolYAsAaPnCYrThAaPnValCYaAPLeuAa	1835
Db	3901	TCCTGCACACTGTGAATCCAGGTTTACATAGGTGACAACTGTACTAATGTGTGTGACCTTGAC	3960
OY	1836	ProCYaSgluHISglInSerValCYeThrArgLYsProSerAlaProHISglYTYrThCYa	1855
Db	3961	CGGTGTAGACACCACTCTGTGTATCCCGCAAGCCCACTGCCCCCAATGCTTAATCTTGC	4020
OY	1856	GlucYAPProProAaPnTYleuGlYProTYrCYaSgluThrArgLIleAaPnProCYaPro	1875
Db	4021	GAGTGTCCCCCAATTACTTGTGGGCATACTGTGAGCCAGAAATTGACCAAGCTTGTCC	4080
OY	1876	ArgglYTPTrPrGlYHISProThrCYaSglYProCYaAaPnCYaAPValSerLYeGlYPhe	1895
Db	4081	CGTGCTGTGTGGGACATCCCAATGTGGCCCAATGACAACTGTGATGTGCAGCAAAAGGCTTT	4140
OY	1896	AaPProAaPnCYaAaPnLYeThrSerGlYgluCYaHISCYeLYsGluAaHISLYrArgPro	1915
Db	4141	GACCCAGACTGCACAAACAAACAAACCGGCGAAGTGTCCACTGTCAAGAGAACCACTACCGGCC	4200
OY	1916	ProGlYSerProThrCYaLeuLeuCYaAaPnCYrTYrProThrGlYSerLeuSerArgVal	1935
Db	4201	CCAGGCAAGCCCCCACTGCTCTGTGTGTACTGTACCCCAAGGCTCTTGTCCAGAGTCC	4260
OY	1936	CysAaPProGlYAsArgGlYInCYeProCYeLYsProGlYValLIleGlYArgInCYaAP	1955
Db	4261	TGTGACCTTGAGAGATGGGCAAGTGTCAATGCAAGCCAGGTGTCAATCGGGCGTCAAGTGTGAC	4320
OY	1956	ArgCYaAaPnProPheAlaGlInValThrThrAaPnGlYCYaGluValAaPnLYrAaPnSer	1975
Db	4321	CGCTGTGACAAACCTTTTGTGAGGTGACCAACAAATGCTGTGAAGTGAATTAATGACAGC	4380
OY	1976	CysProAaPnAlaIleGlInAlaGlYIleTPTrProArgThrArgPheGlYLeuProAla	1995
Db	4381	TGCCCAACAGAGCATTTGAGGCTGTGGATCTGTGGCCCCCTTACCCGCTTCGGGCTGCTGCT	4440
OY	1996	AlaAlaPProCYaPProLYsGlYSerPheGlYThrAlaValAlaAGHISCYaAaPnLHISArg	2015
Db	4441	GCTGCTCCTCTTCCAAAGGCTCTCTTGGAGCTGTGTGGCCCACTGTGATGTGAAGCAACAG	4500
OY	2016	GlYTPLeuProProAaPnLeuPheAaPnCYrThrSerLIleThrPheSerGlInLeuLYsGlY	2035
Db	4501	GGGTGGCTCCCCCAAACTCTTCAACTGACAGTGCATCACTTCTCAAACTGAAGAGGC	4560
OY	2036	PheAlaGlInAGluGlInArgAaPnGlYSerGlYLeuAaPnSerGlYArgSerGlInGlInLeu	2055
Db	4561	TTTCGTGTAGCGGCTAACAGCGGAAATGATCAAGCTTAAGCTCAAGGCGCCTCCCAACACTTA	4620
OY	2056	AlaLeuLeuLeuAArgAaPnAlaThrGlInHISThrAlaGlYTYrPheGlYSerAaPnValLYs	2075
Db	4621	GCCTGCTCTCCGCGCAAGCCACGACGACACCACTGCTACTTCCGAGACGACGTACAG	4680
OY	2076	ValAlaTYrGlInLeuAlaThrArgLeuLeuAlaHISgluSerThrArgInArgGlYpHeGlY	2095
Db	4681	GTTGCGCTTACAGCTGGCCACCGCGCTGTGGCCACAGAGAACCCACGCGGGGCTTTGGG	4740
OY	2096	LeuSerAlaThrArgInAaPnValHISpHeThrGlInSerLeuLeuAArgValISglYSerAlaLeu	2115
Db	4741	CTGTCTGCACACAGAGCTGTCACTTCACTGAAGATCTGTGCGGGGTGGGACGGCCCTTC	4800

[illegible]



Db	5881	ACGGGCCCCAATGGGCTTCTACTACATGCTGGGCTGGGGCCGTGCTGCTTCAATCAAGGG	5940
Qy	2496	LeuAlaValGlyLeuAspProGluGlyTyrGlyValAspProAspPheCysTrpLeuSerIle	2515
Db	5941	CTAGCCCGTGGAGCCCTGGAGCCCGGAGGGCTCAAGGAACTCTGACTTCTGCTCTTCATC	6000
Qy	2516	TyrAspThrLeuIleTyrPsrPheAlaGlyProValAlaPheAlaValSerMetSerVal	2535
Db	6001	TATACACAGCTCATCTGGAGATTGGCTGGCCGGCTGGCTTGGCCGTCTGCATAGATGTC	6060
Qy	2536	PheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluIys	2555
Db	6061	TTCCCTGTACATCCCTGGCGGCCCCGGGCGCTCTGTCTGCTCCAGCGGAGGGCTTTGAGAG	6120
Qy	2556	LysGlyProAlaSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThr	2575
Db	6121	AAAGGTCCTGCTCTGGGCTGGAGCCCTCTTCCCGCTTCTGCTGCTGAGGCCACG	6180
Qy	2576	TrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAla	2595
Db	6181	TGGCTGTGGAGCTGCTCTCTGTCAACAGGAGACCCCTCCCTTCCACTACTCTTTGGCT	6240
Qy	2596	ThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlu	2615
Db	6241	ACCGCAATTGCATCCAGGGCCCCCTTCATCTTCTCTCATATGGTGCTTAGCAAGAGAG	6300
Qy	2616	ValArgIysAlaLeuLysLeuAlaCysSerArgIysPsrProSerProAspProAlaLeuThr	2635
Db	6301	GTCGGAAAGCATCTCAAGCTTGGCTCTGCAGCCGCAAGCCCAAGCCCTGACCTGCTTGAAC	6360
Qy	2636	ThrIysSerThrIleuThrSerSerTyrAsnCyAspProSerProTyrAlaAspGlyArgLeu	2655
Db	6361	ACCAAGTCCACCCCTGACCTCTGCTTACAACTCCCGACCCCTCAAGCAAGTGGCGGCTG	6420
Qy	2656	TyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSaraArgSerGlyLysSer	2675
Db	6421	TACCAGCCCTTAGGAGAGCTCGGCGGCTCTGTGCACAGACCAAGTGGTGGGCAAGAT	6480
Qy	2676	GlnProSerTyrIleProPheLeuLeuArgGlnIleSerAlaLeuAspProGlyGlnGly	2695
Db	6481	CAGCCCAAGCTCATCCCTTCTTGTCTGAGGAGAGTCCGCACTGAACCTCGGCCAAGGG	6540
Qy	2696	ProProGlyLeuGlyAspProGlySerIleuPheLeuGlnGlyGlnAspGlnGlnHisAsp	2715
Db	6541	CCCCCTGGCCCTGGGAGATCCAGGAGAGCTGTCTTCTGAAAGGTCAAGACCAAGCAAGAT	6600
Qy	2716	ProAspThrAspSerAspSerAspLeuSerLeuGlnIleAspArgGlnSerGlySerTyrAla	2735
Db	6601	CTTACACAGCACTCCGACAGTGACTGTCTCTTAAGAACACACAGAGTGGCTCTTAATCC	6660
Qy	2736	SerThrHisSerSerAspSerAspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaPhe	2755
Db	6661	TCTACCCCACTCATGACACATGAGAGAGGAAACAGAGGAGGAGAAAGAGAGAGCGCGCTTC	6720
Qy	2756	ProGlyGlnGlnGlnGlyTTPAspSerLeuLeuGlyProGlyAlaGlyAlaGlyLeuProLeuHis	2775
Db	6721	CTTGAAGAGCAGGGCTGGAGTAAAGCTGTGGGGGCTGGAGCAAGAGACTGCCCCCTGAC	6780
Qy	2776	SerThrProLysAspGlyGlyProGlyProGlyLysAlaProTyrProGlyAspPheGly	2795
Db	6781	AGTACTCCCAAGGATGGGGGCCCCAAGGGCTGTGGCAAGGCCCTCTGGCCAGAGACTTTGGG	6840
Qy	2796	ThrThrAlaLysGlyLysSerSerGlyAsnGlyAlaProGlnGlnAlaGlnLeuArgGlyLysAsnGly	2815
Db	6841	ACCAACAGAAAGAGATGATGGCAACGGGGCCCTTAGAGAGCGAGCTGGGAGAAATGGA	6900
Qy	2816	AspAlaLeuSerAspArgGlnGlySerIleuGlyProLeuProGlySerSerAlaGlnProHis	2835
Db	6901	GATGCCCTGTCTCGAGAGGGGATCCCTAAGGCCCTTTCAGAGCTTTTGGCCCAAGCTCAC	6960
Qy	2836	LysGlyIleLeuLysLysGlyLeuProThrIleSerGlyLysSerSerLeuLeuArg	2855

Db		6961	AAAGGACATCCTTAAGAAGATGTCTCCGCCACCATCATGCAGGAAGAAGCACGCTTCCTGG	7020
Oy		2856	LeuProlenGIuNGInCyThrInglySerSerAArgIySeSerAlaSerGIuNGlySerAArg	2875
Db		7021	CTCCCCCTTGAGAGCAATGCACAGGGTCTTCCCAGGGCTCCCTCGGTATGTAGGGCAGCGG	7080
Oy		2876	GlyGIyProProAaGPProPProAArgGIserIeuNGInGIuNGlyeAuAngLyVal	2895
Db		7081	GGCGGCCCCCTCCCGGCCACCGCCCCGGCAGAGCTCCAGAGACAGCTGAAAGGGGCTC	7140
Oy		2896	MetProIIeaIametsrIlleYalaGIYThValaIpGIuNapSerSergIySerGIu	2915
Db		7141	ATGCCCATCCCATGAGCATCAAGCGCAGCGACGGTGGATGAGACTCGTAGCTCCGAA	7200
Oy		2916	PhelenuPhephaanPheLenuHIS	2923
Db		7201	TTCCTCTTTTAACTTCTCTGCAT	7224
RESULT 8				
LOCUS	AX646739	24370 bp	DNA	linear
DEFINITION	Sequence 931 from Patent Epl270724.			PAT 04-MAR-2003
ACCESSION	AX646739			
VERSION	AX646739.1	GI:28799122		
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	Swu,M., Asai,K., Akiyama,Y. and Aburatani,H.			
JOURNAL	Guanosine triphosphate-binding protein coupled receptors Patent: EP 1270724-A 931 02-JAN-2003; National Institute of Advanced Industrial Science and Technology, Ltd. (JP) ; Center for Advanced Science and Technology Incubation, Ltd. (JP)			
FEATURES				
Source	Location/Qualifiers			
	1..24370			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
CDS	join(201..3510,8553..9200,11163..11385,11634..11749, 11829..12017,12408..12565,12927..13090,13263..13388, 13732..13898,14199..14486,14572..14749,14989..15130, 15312..15432,15856..16039,16226..16327,17669..17763, 17972..18181,18702..18885,19001..19129,19231..19411, 19545..19712,19815..19941,20054..20254,20547..20721, 21048..21160,21338..21417,21507..21633,21721..21843, 22399..22526,22761..22849,22954..23148,23287..23457, 23557..23791,24143..24170) /note="unnamed protein product" /codon_start=1 /protein_id="CAD69495.1" /db_xref="GI:28799123"			
	/translation="MRSAPATGVLPITPPPIALLLLLPPLIGDOVEGCRSLSGR RGSSGACPMGLCPSSASNLMTYTRSCDAGELGHVPHNDGLKWPRESRAHP LPAPRECPMSCLLGICHLSPQKLTLPBHRPCAKAPLRROSKTLAPRLAGE KSTBSFGGRKKRNVTNPFPQPFSTQATVPFETDRKSTHFVFVAQDDMPRSALAT TMALPEPSRNGFSLDPVGAIVATAEIELEKTSKTPVAVIAQDHMPRSALAT ILVTEINDHDHPVEOQRYKSLSLENLEVGVLTVAISTDAIDPNANIIVRLSEGS SPEVEINDIPRSVGIVIRGCVDEEVESVQLVTEASQGDGPPRSTFAVLIFEGSD NDNAPGSRKYVVQVEDYTPGAPVLAIVASRDGNSNVHYVISNGARGFYLD AQCGLDVYSPLIYFTTKETKTFLVRADQGRPLNSGVLYVOVDIDNNAPIPVST PFOATLASPYPLGYLVTHQAIDAIDGNARKLEYRLAGVCHDPPRTINCTGISVAA ELBREEDVSVDFEARDDGTPALNTLASVSIVTVLVNDNNPTFTQEPITYRLMEDNA VGVSVTVSAVDADASHVITYITSGTNRRFSITSQGGJVSIALPIDYKLERQVY LATVASGTRQDAQIVANTVANDTRHPVSOHSHTVNVEDPAQTVALISATED TGNARITTFEMDSIQPRFIDADTGAVTAEILDYEDOVSYTALTARDNGIPOKSDT TYALEILVNDNAPOLFSDSYOSGVDEVPFVSQISATRDGSLNRVYTFEGD GGDGGEFIESTSGIVRTLRLDRNVAVQVLAIVADKGMPARPMPRVATVLDV NDNPVPEODRPVPEVENSPIGLAAVATATPDDETCAOIWQIVQENIPVPOIF TFSELTLATDLDYEDPERVYLQATSAPVVSATATYHALIDNNPPPLGAPEIF NNVYNRSSFFPGAIAGRVAHPDIDSILTYSERENENISLVLANASTELKLSSLRL			



DNRPLEAMSVLVDGHSVTAQCAKVTITIDEMLRHSITLRLSDMSPERFSLPL  
GLFIOAVATLTPPDHVVNVNORDTAPAGHILNVSLVSPQPGPGGPPLESD  
LQERLYANRSLTALTAQVVPVLPFDNI CLRECEMNYCVSLTARSDSAPFASSTVL  
FRPIHVGRLCRCPGFTGDCETEDVLCYSRPGCPHRCGRSGGTCLCRDNYTG  
HCEVSARSGRCTPGVCKNGGTCAVLVNGFKDCPSDPEKPYCOVTRSPDASPT  
TFRGRORPHFTLALSPATKREDDGLLVNGRPNRGRDPALEVOITBOVLTSSAGEST  
TTVSPPVFGVSGSDGMHTVQLKTNKPLLGOTGLFQGSBQKVAIVTDGCTGVALLR  
FESVLGANSYCAAGTQGSKSLDLTGPLLGGVPLDPSPBPVNRORVPGCMRNQVD  
SRHIDMAPIFANGTVPGCPACKVNCSTNCGGTCAVOMDAFCECPGLGSGSCA  
QEMANPOHPLGSSLVAMHGLSLPISQPVLSLMPRTRODGLVALITRGRSTTLQD  
REGHMLSVBEGTLOASSLRLEPGRANDMHAQALQSRVSDTPREGNAILSPYQORA  
EGNGLPRHGLHLSTNITVIGGIPGAGVARGRGLQGRVSDTPREGNAILSPYQORA  
INVEGCSLPDPCDSNPPCANYSCEMNDSTSCDPCGTGNCNVCNLDNPECQSV  
CTKRSAPHCYCECPNLYGCESTRIDQPCRVWMCNPTGCPNCDVSGKGPDDCN  
KTRSGCHCKENHTRPPGSPYCLCDQCYTGSLSRVCDDEDCQCPKPGVIGROCDRC  
NPPAVNTNGCEVNDSCPRALGAIWMPRTFGLPALAPCPKSGVAVHRCDEHSG  
MLPVLNCTSTTFSELGFAERLQNRSGDLSRSQLALLRNATOTHTGREGSDV  
KVAVOLATRLIAHSTORGFLSATODYHPTNLRVNSALDTPKRNHMLIQTEBG  
GTANLQHTYEAASHLQOMRTTYSPTTITNTNVLISVRLDKNRFAGALPRTYAL  
RGEOPDLLETVTLPSVFRETPPVVRPAGPBAQBPBELARORHRELQGEAVAS  
VLIYRTLAGLPHNYDPKRSILRVPKRPITNPVVISVHDBELPALDKPVVPC  
RLLETERTKLCVFMNHSILVSGTGSARGCEVFPNESHVSQCNHMTSPVAVD  
VSRRENGELPLKTLTVVAGTVLALLTPPELTLILSNONGIRBNITLALGLA  
OLVELGINOADLPACTVIALILHPLCTSPWALLRALHLRYLTVRVNTPGPR  
PYVNLGKVPAPITGLANGDPBEGANDPCWLSITDTLINSFACPVAFVMSVFLY  
ILAAASCAQORQGRKGPVSGLOPSFAVILLSATVLLINSSTDLPLHLYFAT  
CNCIOGPIFLSYVLSKEVRKALCLASRKPDPALTTSTLSTNCSPYADGR  
LYOPYDSDAGSLHSTRSKQSPYIPLLEBESALNAGPAGDQMSLFLGQDQ  
OHDPDSDSLSDSDQSGYASTHSDSEBEEBEEBAAPBEOQMSLFLGQDQ  
RLPHSTKDGSGPGRKAPWPGDPTGTTRKSSGNGCAPRERARNDDALSREGSGLP  
GSSAOPHKGLIKKCKLPTISKSSSLRPLEOCTSSSSGNSBSRGRGPPRPPRPO  
SLOBLNKVMPIKMSIKAGTVDESSSEPLFPNPLH

BASE COUNT 4543 a 7241 c 7458 g 5128 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 24370  
Score: 12339.00 Matches: 2907  
Percent Similarity: 36.92% Conservative: 0  
Best Local Similarity: 36.92% Mismatches: 8  
Query Match: 79.38% Indels: 4967  
DB: 6 Gaps: 32

US-09-916-849a-3 (1-2923) x AK646739 (1-24370)

Qy 1 MetatserProialathgylValProleupProthrProproProleuleuleu 20  
Db 201 ATGGGAGCCCGGACCGGGGTCCTCCCAAGCGCGCGCGCTGTGTGTG 260  
Qy 21 LeuleuleuleuleupProproProleuleuglyabglnvalglyProCyahrgserleu 40  
Db 261 TTGCTGCTGCTGCTGCCCGGCACATATGGAGACCAAGTGGGGCCCTGTGCTCTTG 320  
Qy 41 GlyseratrglyargglyserserglyValaCyasaIaPrometgLYTPleuCyapros 60  
Db 321 GGGTCACAGGACGAGGCTCTTCGGGGGCTTCGCCCCCAGTGGGCTCTGTCAATCC 380  
Qy 61 SeralaseramleuTPleuTYrThserArgCyatgaaspalaglyThrgluLeuthr 80  
Db 381 TCAGGGTGAACCTGTGCTCTACACGCGGCTCAGGAGATGGGGGACATGAGTGAAT 440  
Qy 81 GlyhisleuValProhihiaberglyLeuargValITrCyepProgluserGluAhi 100  
Db 441 GGCCACCTGGTGAACCCACACGATGGCTGAGGGTTGTGTGCCAGATCCGAGGCCAT 500  
Qy 101 IleProleupProProalAProgluglyCySProITPserCyatrgleuleuglyIle 120  
Db 501 ATTCCCTTCAACACGAGCTCTGAAGGCTGCCCTGAGCTGTGCTCTTGGGCAATTGGA 560  
Qy 121 GlyhisleuSerProgluglyLeuLeuthrleuProglugluhiIspProCyaleuVala 140  
Db 561 GGCCACCTTCCCAAGGCAAGCTCACATGCGCCGAGGACACCGGTCTTAAAGGCT 620

Qy 141 ProargleuargCyegInserCyaleyleuValaIaIaProglYleuargIaIa 160  
Db 621 CCAAGGCTCAAGTGCAGAGTCTCTGCAAGTGGCAAGGCCCCGGGCTCAGGGCAGGGGA 680  
Qy 161 ArgserProglugluSerleuglyglYargarglyabglnvalaerThralaProgl 180  
Db 681 AGGTTCACAGAAAGTCCCTGGGCGTGGGCGTGGAAAGAAATGTAAATACAGCCCCCAG 740  
Qy 181 PheglInProPserArgYrGlnaIaThrValaProgluansglInProalaGlyThrProva 200  
Db 741 TTCAGCCCCCAGCTACAGGACCAAGTCCGAGAAACAGACCAAGGACCCCTGT 800  
Qy 201 AlaserleuargalIaIaaspProaspIuglygluIaIaIaIaIaIaIaIaIaIa 220  
Db 801 GCATCCCTGAGGGCCATGACCCCGGACGAGGCTGAGGAGGCTGAGGAGTCAAGCAATG 860  
Qy 221 AspalaleuPheaspserArgserArgInPhePheSerleuaspProvalThrglyAla 240  
Db 861 GATGCCCTCTTGTATACCGGCTCCACACAGTCTTCCCTGAGACCCAGTCACTGTGCA 920  
Qy 241 ValThrThralaIa 260  
Db 921 GTACACACGCGGAGAGCTGATCTGTGACCAAGACCAACCCGCTTCAAGGCTCAG 980  
Qy 261 AlaglnasphIsgIymetProargargseralaleuValaIaIaIaIaIaIaIaIaIa 280  
Db 981 GCGCAGAGCAACGCGCAGTCCCGGACGAAAGTCCCTGGCTACCTCACTCACTTGGTACT 1040  
Qy 281 AsphtrhmasphIasapProvalaPhegluglnIuglnIuglyglYleuLeuargIu 300  
Db 1041 GACACCAATGACCAATGACCTGTGTGTGAGCAGGAGGACCAAGAGAGCTCAGGAG 1100  
Qy 301 AsnleuIuValgIyYrGlnuValleuThrValaIaIaIaIaIaIaIaIaIaIaIaIa 320  
Db 1101 AACCTGAGGTGCTGATGAGTCTCACTGTGAGGACCAAGATGATGATGATGATGATGAT 1160  
Qy 321 AsnIa 340  
Db 1161 AATCCCAATATTCGTGACCCGCTGTGAGAGGCTGTGGGGCAGACCCCTCGAAGTCTTT 1220  
Qy 341 GluIleaspproaspserglValIleargThrargIyProvalaIaIaIaIaIaIaIaIaIa 360  
Db 1221 GAGATCAACCTGCTGTGGGGTATCCGAACCCGTGACCTGTGATCGGAAAGAGGTG 1280  
Qy 361 GluSerYrGlnuLeuthrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 380  
Db 1281 GAATCTTACAGAGTGAAGTGAAGGCAAGTGAACAGAGGTGGGACCCGGGTCTCGAGAT 1340  
Qy 381 ThrThralaIa 400  
Db 1341 ACCACACCCGCTGTCTTCTGTGTGAGAGATGACCAATATATATGCCCCCAGATTAGT 1400  
Qy 401 GluIuYsaIrgYrValaIa 420  
Db 1401 GAGAAAGCTATATGTGCTCAGGTGAAGAGATGTACTCAGGGGCCCAAGTACTCGGA 1460  
Qy 421 ValThralaseraspargaplyglYserAsnAlaValaIaIaIaIaIaIaIaIaIaIa 440  
Db 1461 GTCAACGCTCGATGAGCAAGGAGGAGCAATGCGTGTGCACTATGATGATGATGATGAT 1520  
Qy 441 GlyasnaIa 460  
Db 1521 GGCATATCTGGGGGACAGTTTATCTGAGTCCGACGCTGAGCTGTGATGTGTGAGC 1580  
Qy 461 ProleuaspYrGlnuThrThrlygluYrThrleuargValaIaIaIaIaIaIaIaIaIa 480  
Db 1581 CTTCTTACATGAGAGCAACAGAGTACACCTTACGGGTGCAAGCAAGATGATGATGATGAT 1640  
Qy 481 ArgProProleuSerAsnAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 500  
Db 1641 CGTCCCACTCTATATGTCTGTGGCTGTGACAGTACAGGCTCGTGAATATCAAGAC 1700  
Qy 501 AsnIa 520

Db	1701	AATGCCCCCACTTCCTCGTAGCAACCCCTTCCAGGCTACTGCTCGTAGAGGTGCCCTTA	1760
Oy	521	GIYTYrLeuValIleuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	1761	GGCTAACCTGGTCTCCAGTGTCCAGGCTATGACGGTATGTGTGGAGCAAAAGCCGCTCG	1820
Oy	541	GIuTYrArgLeuAlaGlyValGlyYHisAspPheProPheTrpIleAsnAsnGlyTYHcIy	560
Db	1821	GAATACCGCCTTGCTGGGGTGGGACATACCTTCCCTTCAACATCAACAAATGGACAGGC	1880
Oy	561	TrpIleSerValAlaAlaGluLeuAspArgGluValAspPheTYrSerPheGlyVal	580
Db	1881	TGGATCTCTGGGGCTGTGAACCTGGACCGGAGGAAAGTTGATTCTACAGCTTTGGGGTA	1940
Oy	581	GIuAlaArgAspHisSGLYTHrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
Db	1941	GAAGCTCGAGACCAATGGCACTTCACGACTTACCTGCTGGCCAGTGTACGGGTACGTGC	2000
Oy	601	LeuAspValaAsnAspAsnAsnProThrPheThrGlnProGluTYrThrValArgLeuAsn	620
Db	2001	CTGAGATGCAACGACCAATCCAACTTTATCCCAACAGAGTACAGAGTACAGTGGCTCAAT	2060
Oy	621	GluAspAlaAlaValGlyTHrSerValValThrValSerAlaValAspArgAspAlaHis	640
Db	2061	GAGATGACGGTGGGGCCACCAAGGTGTGACGGGTGTCAGGTGTGACCCGTGAATGCTCAT	2120
Oy	641	SerValIleThrTYrGlnIleTHrSerGlyAsnThrArgAsnArgPheSerIleTHrSer	660
Db	2121	AGTGTCACTCACTTACCAAGATCAACAGTGGCAATCTGCAAAACCGCTTCCATCAACCAAC	2180
Oy	661	GlnSerGlyGlyGlyLeuValSerIleuAlaLeuProLeuAspTYrLYbLeuGluArgGln	680
Db	2181	CAAGGTGTGTGGGTGGTGTATCCCTTGGCCCTGCACTGAGCTTCAAACTTGACGGCGAG	2240
Oy	681	TYrValIleuAlaValAlThrAlaSerAspGlyThrArgGluAspTrpAlaGlnIleValVal	700
Db	2241	TATGTGTGGCTGTTAACCGCTCCGATGTGCACTGGCAGGACAGGCAACAATGTGGTG	2300
Oy	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYrThrValAsn	720
Db	2301	AATGTCAACCGACGCCCAACCAACCATGTCTGTCTTCAGACTCCCACTTATACAGTGAAT	2360
Oy	721	ValaAsnGluAspArgProAlaGlyTHrThrValValIleuIleSerAlaThrAspGluAsp	740
Db	2361	GTTATATGAGACCGGGCCGAGGACCAACGGGTGGTGAATCAAGAGGCCAAGATAGAGAC	2420
Oy	741	ThrGlyGluAsnAlaArgIleTHrTYrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2421	ACAAGTGAAGATGCCCGCATCACTTCACTTACAGAGACACACATCCCACTTCCGATC	2480
Oy	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTYrGluAspGlnValSer	780
Db	2481	GATCGACGACAGGGGGGTGTCAACCAACGAGCTGAGCTGACATATGAAAGCAAGTGTCT	2540
Oy	781	TYrThrIleuAlaIleTHrAlaArgAspAsnGlyIleProGlnLYbSerAspThrThrTYr	800
Db	2541	THACACCTGGCCATTACTGTCTCGGGACATGGCATTTCCCGAAGTCCGACACCACTTAC	2600
Oy	801	LeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerTYr	820
Db	2601	CTGGAGATCCCGGTGAGACGATGAATGACATAGCCCTCAAGTTCCTCGAGACTCCAC	2660
Oy	821	GlnGlySerValTYrGluAspValProProPheThrSerValIleuGlnIleSerAlaThr	840
Db	2661	CAGGGCAATGTCTATAGAGATGTGCCACCTTCACTACAGTCTCGCAATCTCAACCACT	2720
Oy	841	AspArgAspSerGlyLeuAsnGlyArgValPheTYrThrPheGlnGlyGlyAspAspGly	860
Db	2721	GATCGTGAATTCAGACTTATATGGACGGGTCTTCTTCACTTCAAGAGAGGGACGATGGA	2780
Oy	861	AspGlyAspPheIleValGluSerThrSerGlyIleValaGlnThrLeuArgArgLeuAsp	880

D	2781	GAGGCTACCTTAA	TTGTTAGTCCAGCTCAGGCACTGGGAACGTAACGAGGCTGGAT	2840	
Q	881	ArgGlu	asnValAla	glnTyrValLeuArgAlaTyrAlaValAspIleGlyMetProPro	900
D	2841	CGAAGAAACGTGGCC	AGTATGCTTTCGCGGCAATATGCAATGGAACAAGGGATGCCCCCA	2900	
Q	901	AlaArgThrPro	MetGluValThrValThrValLeuAspValAsnAspAsnProProVal	920	
D	2901	GCCCGCACCTTA	TAGAAAGTGAACGTACCTGTGTGATGTGAATGACATCCCCCTGTC	2960	
Q	921	PheGluGlnAspGlu	PheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal	940	
D	2961	TTTATAGACAGATGA	GTGATGTGTGTGTGGAAAGAAACAGCCCATTTGGGCTGACCGGTG	3020	
Q	941	AlaArgValThr	AlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle	960	
D	3021	GCCCGGGCTACAG	CCACTGAACCCCGATGAAGGACCAATGCCAGATTAATGTAACAATT	3080	
Q	961	ValGluGlyAsn	IleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla	980	
D	3081	GTGAGAGGCAACAT	CCCTGAAGTCTTTCACTGGACATCTTCTCCGGGAGCTGACAGCC	3140	
Q	981	LeuValAspLeu	AspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer	1000	
D	3141	CTGTGTAACTTGA	CTACGAGAACCCGGCTGAGTACCTCTTGATCAACAGGCCACGTCA	3200	
Q	1001	AlaProLeuValSer	ArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020	
D	3201	GCTCCTCTGTGTAG	CCGGGCTTACGTCCACTCCGCTCTTGAACCGCATGACAAACCA	3260	
Q	1021	ProValLeuGly	AsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040	
D	3261	CCAGTGTCTGGGCA	ACTTGGATCTCTTTCACAACAACATATGTCAACATCTCCTCAAGAGC	3320	
Q	1041	PheProGlyGly	AlaIleGlyArgValProAlaHisAspProAspIleSerArgSerLeu	1060	
D	3321	TTCCCTGGGGGTG	CCATTGGCCAGATACCTGATATCTCAGATATGCTG	3380	
Q	1061	ThrTyrSerPhe	GluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080	
D	3381	ACTTACAGCTTTTG	AGCGGGAAATGAATCAGCTGGTCTCTGCTCAATGTCTCAACGGGT	3440	
Q	1081	GluLeuLeuSer	LeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal	1100	
D	3441	GAGGTGAAGCTAA	GCAGCGCACTGGAACAACAACCGGCTCTTGAGAGCCATCATATGAGCTG	3500	
Q	1101	LeuValSer		1103	
D	3501	CTGGTGTTC	AGGTAAAGGAAGGCCACAGTGGCCGTGGGGGTATGCTCGCGGGAGT	3559	
Q	1103			1103	
D	3560	GTCTGGGACAG	CACTGAGAGTAAAGTGCATCCAGAAAGACAGTACAGATCCACTTCCT	3619	
Q	1103			1103	
D	3620	GCCCAAGTCTG	GCACAGACAGAGAGGGGCAAGACGACTTGGGAAGAAAGCCCAAGAA	3679	
Q	1103			1103	
D	3680	CCCTGGCGGCTG	TGACAGGCCCACTTCTGCTGGGCTGAGGCTGGCTTCGAGTTGGC	3739	
Q	1103			1103	
D	3740	TGCCCCACTGT	CTACAGCCTGTGACGTGTGTGGGGAAGTGTGTGAGGCTGCCCGGC	3799	
Q	1103			1103	
D	3800	TATGCTGCAAGA	ATTGTGAAAAAAGCGGTGTGAAGCTTGCGGGGGCCAGAGAGCTGG	3859	
Q	1103			1103	
D	3860	GGAAGAGGCA	GAGAGCACTGTGGCGCAGAGAGGAGCGAGCCAGGAGCCCTTCAACCC	3919	

QY	1103	-----	1103	-----	1103
Db	3920	CATCAAGATATTCTCTGAGTCTCACACAGCCTTTGGCAGAGCTGGGGACACAGGGAGAC	3979		
QY	1103	-----	1103	-----	1103
Db	3980	AGGGCTGCCCAACCCAGGGGTGAGAACCTGTAGTCTCTTGTGGCCCTTTTGCTCC	4039		
QY	1103	-----	1103	-----	1103
Db	4040	TGGCTGAGCTGCAAGATCGGCTGACAGGACCTTTGTAGCGCATTTCTTCCCTCCAGG	4099		
QY	1103	-----	1103	-----	1103
Db	4100	CAGAAAGAGCTGGCTGGCTCCCTTCAAGAGCTCCACAGAAAGTGAGGCGAGAGCTC	4159		
QY	1103	-----	1103	-----	1103
Db	4160	ATTGTCTGTCCAAACAGACCCCATTTGAGATCGGTCTGGGTCTGGCTGGAGAG	4219		
QY	1103	-----	1103	-----	1103
Db	4220	GGGCTGCAATGAGCTGGGGGGCAGGGCTGAGGCTCTGGGCCCCCTTGCCACTCATG	4279		
QY	1103	-----	1103	-----	1103
Db	4280	GGCTGTCTTAGGGTCCAGGCTTCTTCTCTGCTAGCTGTGACAGGCCAAGCTCAG	4339		
QY	1103	-----	1103	-----	1103
Db	4340	GAATGAGTGTAGGACCAAGACAGGCACTGTGACTGTGTAGCTTTGAGGCGACAGACC	4399		
QY	1103	-----	1103	-----	1103
Db	4400	CAAGGCACTTATCTGAGAGAGAAAGAGTCCAGGGCGTGGAGCTGGCCGGAAACC	4459		
QY	1103	-----	1103	-----	1103
Db	4460	AGGCTGGGCAATTCCTGCTGCTTGGCCAAAGCATCAGCCTACAGATCCCTGACAGAG	4519		
QY	1103	-----	1103	-----	1103
Db	4520	GGCTGGGGAGAGATTGGAGGGGTTGTGACAGTATCTGTCTTTTATCAAGCTG	4579		
QY	1103	-----	1103	-----	1103
Db	4580	TGGGAGCGCAGAGGTAGCAGCAGGAGAAACACACCCAAAGTCTGAGTGCCTTCCC	4639		
QY	1103	-----	1103	-----	1103
Db	4640	AAGCTGGGAAACAATGAGACCTCATATGAGGGAGTGTGTAGCTGCACCCAGGGGTGG	4699		
QY	1103	-----	1103	-----	1103
Db	4700	ATTGATGCATTCCTCTTAGGCAAGCTTCAATTCATCAGGAAAGGAGAAAGAGA	4759		
QY	1103	-----	1103	-----	1103
Db	4760	GGCTGGCAGGCCCTAATCCACCCAGGAGTGTGGCTGAAAGGAGAGGGGTATG	4819		
QY	1103	-----	1103	-----	1103
Db	4820	CCAGAGTCAGGGCGTGGCTCATCTCAGGGTCTTGTGACAGATGCTCCCTGCCCCAC	4879		
QY	1103	-----	1103	-----	1103
Db	4880	CCTTCAAGTATGCCACAGGAAACAGGAAAGCAGAGCTTTTCAAGGACTGTGGCTGAGTT	4939		
QY	1103	-----	1103	-----	1103
Db	4940	CTGAGCTCTGAGACTTCAAGTTCCCTCCAGAGAGCGAAAGAGGAGGTGAGCTCC	4999		

---

QY	1103	-----	1103	-----	1103
Db	5000	TGTGTCTCAATCTGTGCTGGGCACTTGAGACAAAGTGGAAAGAGTGGAAAGAGGGCC	5059		
QY	1103	-----	1103	-----	1103
Db	5060	TCTCTCTGTGCTCCACACCCAGAGCTGCTTGTGAGCGTGGCCAGGCACTGGCAGA	5119		
QY	1103	-----	1103	-----	1103
Db	5120	GAAAGCCCCCAATTAGCCTTGGCTCCACCAAGGCAAGCCCTAGACTTGTGTTCC	5179		
QY	1103	-----	1103	-----	1103
Db	5180	TAGTCTAGAGAAGCTGGGGCAACATCAGAACGGCTCCTCTGGGCCGGCAGCTGTG	5239		
QY	1103	-----	1103	-----	1103
Db	5240	CCCCACCTCTCCTTCTGGCTTGCACCTTGTAGAGGAGTCAACAGCAACCCAAAGTT	5299		
QY	1103	-----	1103	-----	1103
Db	5300	CAGGAGGCGGGGTGAGGGGAGAGCTGATCCACTGCCCCCTTGTCTTCTCTTGT	5359		
QY	1103	-----	1103	-----	1103
Db	5360	TCTTGTCTCTTCTTCTTCCGCCCCCTGCGAGAGGAGCCCTGACAGACTGCTCTC	5419		
QY	1103	-----	1103	-----	1103
Db	5420	TCCGTGCCTCTCCCTGCTCTGCGGCTCTCCGGCTCTCCACTTGCTGGCT	5479		
QY	1103	-----	1103	-----	1103
Db	5480	CTTCTCTGAGCTTGTCTTCTCCCTGTTGTGCTCGTGTGTGCTTCACTCTCTTC	5539		
QY	1103	-----	1103	-----	1103
Db	5540	CCTATGGCTGACAGCTGCCAGCCTGGGCGGATGGCTCCCTGAGAGCTGAATTCCTC	5599		
QY	1103	-----	1103	-----	1103
Db	5600	GCTCTTTTCTTCTCTGCTCAGACAGCAGAGCTCTGTCCATGTCTGGGAGCTCCC	5659		
QY	1103	-----	1103	-----	1103
Db	5660	AGAACAGAGAGCTCTTGGCTCCAGTCTCCGCTGGACATCTGCCAGGGAGAGAG	5719		
QY	1103	-----	1103	-----	1103
Db	5720	GCTGGAGGCTGACCGGAGCTGGCACTTTTCCAGGTGTGCTGTAAAGCTTGGGACAC	5779		
QY	1103	-----	1103	-----	1103
Db	5780	AGGTGCATTTCTTGGGTCAGAGGCCATCCAGCTGTCTCTTCTGTCCAGAAAGACA	5839		
QY	1103	-----	1103	-----	1103
Db	5840	GCTCAGGCAACGTACTCCACCCCTGTTCATCTGTCTTGTACCTGAGGGGTGACC	5899		
QY	1103	-----	1103	-----	1103
Db	5900	TAGTCAAGGCAAGGAGGGGTATAGAAAGAGAGACAGTGGGTGACTGTGGGGA	5959		
QY	1103	-----	1103	-----	1103
Db	5960	TGGGGGACACAGCAGAGGGGGGAGGTGCCACAGATGTGACCAAGTGGAAAGACA	6019		
QY	1103	-----	1103	-----	1103
Db	6020	GAGACATGTGGGAGAGGGGCCAAGCGGGGCTAAGACGATGGCGGTGGGGACACAGG	6079		
QY	1103	-----	1103	-----	1103

Db 6080 GAGAGTGTGTTCCACGGCTGAGGGGTGCAAGTTCTCCCGCTCCAGGCTTGG 6139  
 QY 1103 ----- 1103  
 Db 6140 TTGCCCCCTCTCTCTTAATGATTATGAGGCACTGTGTGAGTCTTGGGCTGTTT 6199  
 QY 1103 ----- 1103  
 Db 6200 CTGCTGAGGGAATATCCGTCAGGAATCATTAGGCAATATACAGGTATACAGAG 6259  
 QY 1103 ----- 1103  
 Db 6260 CTCTGTGCGAGTGAAGATCTGTGCGGGAAGTGTGAACGAGAGGAAGTAGAGCC 6319  
 QY 1103 ----- 1103  
 Db 6320 TGAATGACATTAATCTGGGCAAGGCTTCTGTGAGCAGGTAGCTTAGGGAGTTAGAAAGCT 6379  
 QY 1103 ----- 1103  
 Db 6380 CAGAGACTGTGGGCTGTGTCTGGGTGGAGAGAAACGCTGTGACAGAGCTTTGAGAC 6439  
 QY 1103 ----- 1103  
 Db 6440 GAGAGCAGAGGAACAGAGTGGGTGATGACAGAGGCTTGTGTGAGAGCTAGTAG 6499  
 QY 1103 ----- 1103  
 Db 6500 AATTAAGCAATGATGAGGCTGAGGGTAGGTGGGCACTCGGCTGAGAGACATGGCTGAGAT 6559  
 QY 1103 ----- 1103  
 Db 6560 GGGGGTCTGCTTCAAGCTTCCGGGCTTCACTGAGCACCGGTGATCAGTCAAGCAC 6619  
 QY 1103 ----- 1103  
 Db 6620 GCAAGTAGTACACACAGATGCAAAATTTGTTCCATGTCTTGTCTTCCACCCC 6679  
 QY 1103 ----- 1103  
 Db 6680 ACCCCAGAAACATGGGCTCTGCCCGTACTGTAGGGTAAACATTTGCTTTTATATC 6739  
 QY 1103 ----- 1103  
 Db 6740 TAGTGGGCTCCCCCTGCTATGTGTTTGTGTTGTTGGAGACAGTCACTCTG 6799  
 QY 1103 ----- 1103  
 Db 6800 TCAACCAAGCTGAGTACAAATGGCGGATCTCGGCTCACTGCAACCTCTGCTTGGGT 6859  
 QY 1103 ----- 1103  
 Db 6860 TCAAGCAATCTCTGCTCAAGCTCTGTAGTATCTGGGACCTCTGAGTACGTGAGATT 6919  
 QY 1103 ----- 1103  
 Db 6920 ACAGGCAACCAACACAGGCCAACTAATTTTGTATTTTATAGTAGAGAGGGGTTTAC 6979  
 QY 1103 ----- 1103  
 Db 6980 CATGTGTGATGTCTGAACTCTGAACTCTGATTCACCGGCTCAAGCTTCCCA 7039  
 QY 1103 ----- 1103  
 Db 7040 AAGTTCTGGGACTACAGGCGTAGAGCACTGGGCCGGCCCCCTGTATGTTTAAAGTCC 7099  
 QY 1103 ----- 1103  
 Db 7100 GTTATTTCTTGTGAGACTGAATGATGTCTATCCATCTCAGTGTTTTCTGAACACC 7159  
 QY 1103 ----- 1103

Db 7160 TACTATGTATAGGACTTGTATTAATACAAATATGATATGACATGATCCCTGT 7219  
 QY 1103 ----- 1103  
 Db 7220 AATCCAGTGTCTTAGAAGGCTGAGATGAGAGATCATTTAGGCCAGAAATTTGAGACA 7279  
 QY 1103 ----- 1103  
 Db 7280 GCTTGACACACAGCAGACCCCATCTCTACACAAACAAATATGACGAGCATAGT 7339  
 QY 1103 ----- 1103  
 Db 7340 TGTGACACTGTGTCTCACTAATTTGAGAGTTGAGGTGAGAGATCGCTTAGCCCA 7399  
 QY 1103 ----- 1103  
 Db 7400 GGAGTTCAAGCTACAGTAGATGATGTGCCACTGCTCTCCAGTATAGCCGACAGAG 7459  
 QY 1103 ----- 1103  
 Db 7460 AGTGTCTCTTAAAAAAGATACTATCCCTCTGTGCTGTAAAGCATTT 7519  
 QY 1103 ----- 1103  
 Db 7520 CCTCTTCCAGGCCCGAGCTTTCATTTTATGCTAATAATATTCATCCATGTGTA 7579  
 QY 1103 ----- 1103  
 Db 7580 TGTGATTCAGTGTGTATGTATTTCTGGGCTTAGGGTGTGTGGTGAATGAGACAC 7639  
 QY 1103 ----- 1103  
 Db 7640 AGCAATGTTTATGCTTAGCTGTTAATAGTAGTGTGCTGTGGGGCATAGCA 7699  
 QY 1103 ----- 1103  
 Db 7700 GGCATTTGGGTGTAAATGTGACCAAGTGGTATGATGACTAGAGCCCTAGAGCA 7759  
 QY 1103 ----- 1103  
 Db 7760 GAGAAAGACTGACTCAGTGGGGCTGTGGGGTCTCAGAGTGAAGAGCTGCCAGCCCTT 7819  
 QY 1103 ----- 1103  
 Db 7820 CTATCTGTGCTTCCGTCGCCCAACCCGACTCACCTGTCTCAGCATGGGCAGTGG 7879  
 QY 1103 ----- 1103  
 Db 7880 TGATGATGCTACTGGTTAACGTGCTCAGCTGATTTCTGGGGTGGGGCCCCCAGGCT 7939  
 QY 1103 ----- 1103  
 Db 7940 ACCCAGAGCCGAGAGATTGGCGGGAAGATCTTGTGTGATGTTGCATGGG 7999  
 QY 1103 ----- 1103  
 Db 8000 GATAGAGTCTGCGGCCCAAGCAGCAATGGGCTCAGTGGAGGGGGACTAGGTGCGC 8059  
 QY 1103 ----- 1103  
 Db 8060 CACTGACCAATTTACCACTCCCCCAACCCCGCACTCTGCTGTGTGGGGAG 8119  
 QY 1103 ----- 1103  
 Db 8120 TGGGAAACAGCAGTGGTGGATTATCGTGTGTGAGGTGAGAGGAGGGGCA 8179  
 QY 1103 ----- 1103  
 Db 8180 GGGGATGATTTTGTGGGTCTCAGGATGCTGTGGCCCTGTAGAAAATGATGAAAAAGA 8239  
 QY 1103 ----- 1103  
 Db 8240 CAACATAGGTTCCCGCCAGGACAGACAGACTCTGGGGGAAGGAGACAGAGCGTGGC 8299

QY	1103	-----	1103	QY	1319	-----	1319
Db	8300	TTGGCAGGGAAGGTGCA	8359	Db	9380	GGAGTCTTCCCTGCTGTG	9439
QY	1103	-----	1103	QY	1319	-----	1319
Db	8360	TTGGGAAGATGCAAGAA	8419	Db	9440	CTCCAGACAGTGGCGCT	9499
QY	1103	-----	1103	QY	1319	-----	1319
Db	8420	GGCCACAGTGA	8479	Db	9500	TGAACAAGTTACTTAAC	9559
QY	1103	-----	1103	QY	1319	-----	1319
Db	8480	TGGTGGTGTGTCAGGAA	8539	Db	9560	TTAGGTCAATAGTTGTT	9619
QY	1104	-----	1115	QY	1319	-----	1319
Db	8540	TTCCCTTCCACAGAGCG	8599	Db	9620	TGGCATGGAAGTGTGG	9679
QY	1120	-----	1139	QY	1319	-----	1319
Db	8600	ATCACCAGATGAGATGCT	8659	Db	9680	GTTCTATCACA	9739
QY	1140	-----	1159	QY	1319	-----	1319
Db	8660	ATCACCAGATGAGATGCT	8719	Db	9740	AGCGATTCCTTTCTGA	9799
QY	1160	-----	1179	QY	1319	-----	1319
Db	8720	CCACCGGACACAGTGTG	8779	Db	9800	ACAGAGAAGCTGGGGA	9859
QY	1180	-----	1199	QY	1319	-----	1319
Db	8780	ATCTTCAACGTAGACCT	8839	Db	9860	ATTGCACTGTATGGGCT	9919
QY	1200	-----	1219	QY	1319	-----	1319
Db	8840	CTGCCCTCTGAGGACCT	8899	Db	9920	GGATCTGGCTAGGCGCT	9979
QY	1220	-----	1239	QY	1319	-----	1319
Db	8900	TCGGCAACGCGGCTGCT	8959	Db	9980	CCCTGGATGCTACTGT	10039
QY	1240	-----	1259	QY	1319	-----	1319
Db	8960	TTACATGCGCTGCTGCT	9019	Db	10040	CTTCTCAGACACCCCT	10099
QY	1260	-----	1279	QY	1319	-----	1319
Db	9020	TCGGTCTCTTCCGCGCC	9079	Db	10100	CCTGCTCTTGTCTTCA	10159
QY	1280	-----	1299	QY	1319	-----	1319
Db	9080	TTACAGGCTGATCTACT	9139	Db	10160	CATCTTCCCGCTGAT	10219
QY	1300	-----	1319	QY	1319	-----	1319
Db	9140	CACGGGCGCTGCGACG	9199	Db	10220	CTGCTAAGGGTGGGGG	10279
QY	1319	-----	1319	QY	1319	-----	1319
Db	9200	GGTGAGCCAAAGGAGGA	9259	Db	10280	CAGAGATGTGTGGGGG	10339
QY	1319	-----	1319	QY	1319	-----	1319
Db	9260	ACAAATCAGGACAATGT	9319	Db	10340	ATGAAACAAAGAGCA	10399
QY	1319	-----	1319	QY	1319	-----	1319
Db	9320	CTGAGAGGTCAATTAAT	9379	Db	10400	ATGGGAGGAGAGAGG	10459

```

Db      10460 TGAATTTCTGTCTCTCTCAGAGGCTCAAGTTTCCACTCTCAGCATGTCCAGGTGCACAG 10519
QY      1319 -----
Db      10520 AGGTCTTAGTCTGACCGCAGAGGCTCAAGGTGGAGTCTCTGCCCCATTCTGGGGGC 10579
QY      1319 -----
Db      10580 TCCCGAGTTCTCTCTCATACTCTTCATACCCAGACAAATCTAAGTGAAGCCTGATC 10639
QY      1319 -----
Db      10640 CCATCCCGAGGAGGAGAACTGATCTCAGTGCCTGAGAGGACAAAGACCAAGTGGAA 10699
QY      1319 -----
Db      10700 GGAAGGAGGAGGAGGAGGCTCCAGCCCTCCCGCTGTCTGCAAGGCCCATCTGTTGTGAC 10759
QY      1319 -----
Db      10760 ATTACCTCTTGGGGGCGAGGACACCCAGGTGTGTTGGTAGTGTGACATCGAAGGC 10819
QY      1319 -----
Db      10820 TCCCGGATCCGGGCTTAAACAAGTGTCCAGCAGGCTGTGCTTAATGAACCCAA 10879
QY      1319 -----
Db      10880 TGGGGTAGGGAAGGTCTGGGAGCTCCAGGAATTGCAAGGAGACAGTGCCTGAGACTG 10939
QY      1319 -----
Db      10940 TGGCAGAGCGGCTCAGTGAACCAACAGGTGCCATGAGAGGCTGGGACCATAGATC 10999
QY      1319 -----
Db      11000 TCAGAGAGATGACACACAGAGAGAGTGGGATGCTGGAGGTTTGGGAGTATTACG 11059
QY      1319 -----
Db      11060 AGGTCAAGGAGGCGCATGGAGCTATCATGAGTTTGGGGGCTCAGTTCCCTCTCT 11119
QY      1320 -----
Db      11120 GTCTCAGGTACTGTGTACTCACTGCTTCTTCTTGTCCAGGTGAGCATGTGAGGTG 11179
QY      1326 SerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGlyVal 1325
Db      11180 AGTGTCTGCTCAGGCGGTGACCCCGGGTGTGCAAGATGGGGGCACTGTGTCAAC 11239
QY      1346 LeuLeuValGlyPheLeuGlyCysAspCysProSerGlyAspPheGlyIleValProGlyCys 1365
Db      11240 CTGCTGTGTGGGGGCTTCAAGTGCATTCCTCATCTGGAGACTTGGAGAGGCTCAATGCT 11299
QY      1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGln 1385
Db      11300 CAGGTGACCAACCGCAGCTTCCCCCGCCACTCTTCATCACTTCGCGGCTTGGCCAG 11359
QY      1386 ArgPheHisPheThrLeuAlaLeuSer----- 1394
Db      11360 CGTTTCCACTTCACTCCCTGGCCCTCTC-GTAGATGGCTGGGCACTGGGGGTGGGAGTGGG 11418
QY      1394 -----
Db      11419 CCTGTGGGCACTAGAGTCTCTCTGTCTGTGTGTTAAGTAAGAGTCAAGCAGTGA 11478
QY      1394 -----
Db      11479 AGCTGAGACAGGCTGCCGAGACCAAGTCTGCTCTTGAGATTGCTGTGGTTCTGCGC 11538
QY      1394 -----

```

```

Db      11539 AGCAGATCTCCCTCCCTCGCGTGGTTAAGTGGCGGGGTGTGCAATCCAGGTGGTAC 11598
QY      1395 -----
Db      11599 CCATTCCTGCCCCCATCCCAACTCTCTGTTCAGGTTTGCACAAAGAGCGGAGCGG 11658
QY      1403 LeuLeuValArgPheAsnGlyValArgPheAsnGlyValHisAspPheValAlaLeuGlyValIle 1422
Db      11659 TTGCTGTGTACATAGGCGGTTTCAATGAGAACATGACTTGTGTGCGCTCGAGGTATC 11718
QY      1423 GlnGlnValGlnLeuThrPheSerAla----- 1432
Db      11719 CAGAGAGGTCCAGCTCACCTCTCTGC-AGGTATCAGATTGCCCCCATCTTGCC 11777
QY      1432 -----
Db      11778 CATCTCCAAAGGCCCAAGCTTTCAGCCCTGACCCCAAGCCACATCTATCAGCC 11837
QY      1432 -----
Db      11838 AATCTGGGCCAGCCCAAGCACTGGACCCCAACCTGACATCTAGCCTGCTCA 11897
QY      1433 -----
Db      11898 GGCATTCAGCTCACTGTCTCTTCCCGAGGAGTCAACACAGGTGTCCCATTC 11957
QY      1443 ValProGlyValIleSerAspGlyValIleThrPheThrValGlnLeuValArgPheVal 1462
Db      11958 GTGCCGAGAGATCAATGATGCGCAATGCGTCACTGCGTCAAGTAAATCAATTAAG 12017
QY      1462 -----
Db      12018 GTGGGTGTGAGGACACAGAGGTTGGGGTCTGTCTTGTCTCAGGTCTTACCAAGC 12077
QY      1462 -----
Db      12078 CTTGATGCGCATTTGCTTCCAGGCTTGGGTGGCTGTCAAGGCAATTTGCGTGAAGGA 12137
QY      1462 -----
Db      12138 AGGAGGCGGTGGGAAGTGCACCTTGTGGGCAATCCACCCACATGGGCACTGCTG 12197
QY      1462 -----
Db      12198 ATGCTCTGTTCATAGGACCGCGAATTGGAACATGAGAGATGATTACATGA 12257
QY      1462 -----
Db      12258 TAAACAGTGAAGCACAGTGTGTATATGACATGTGTATGCTGTGGGCAATGG 12317
QY      1462 -----
Db      12318 GTGACACAGAGGACCCAGATGTGTGCTGGCANTGAACCTAGTGGGCTGTACTAGTG 12377
QY      1463 -----
Db      12378 TCCCTCTCTCTGCTCTTCTCTGTCCAGGCACTGTGGGTGAGAGGCTCCACAG 12437
QY      1473 GlnProSerGlnGlnValAlaValAlaValThrValAspGlyCysAspThrGlyValAla 1492
Db      12438 GGGCCATCAGAGCAAGAGGTGCTGTGTGATCCGTGATGCTGTGACACAGAGTGGCC 12497
QY      1493 LeuArgPheGlySerValLeuGlyAsnValSerCysAlaAlaGlnGlyThrGlnGlyVal 1512
Db      12498 TTGCGCTTCGATCTGTCTGTGGCACTACTCTGTCTGCCAGGGCACCCAGGGTGGC 12557
QY      1513 SerValVal----- 1515
Db      12558 AGCAAGAGTGAAGAGGAGAAAGGCCAGGATGGGTGAATGAGAGGCAAGGCCAGGCA 12617
QY      1515 -----
Db      12618 GGAAGGAGGTGAAGAGGGGCTTTTGGGATCTCTCTCCCTTGTGCTCTTACGCC 12677

```



QY 1515 ----- 1515  
 Db 12678 CTGCTCAGCCCTGGGAGATGGGGTCAAGACTGGGGAATCAAGACAAGGCTGTGGGGCAG 12737  
 QY 1515 ----- 1515  
 Db 12738 TCTGGGAGGGGGCTGAGTGTTCAGACAAGAACATGTTCCAGGTAGGGGTGACGCGCTTCT 12797  
 QY 1515 ----- 1515  
 Db 12798 GAGCACTGCTCTGAGAGGGCTCTTGGACACTGGCTACGCGGTTCTGTCTCTGGCTGG 12857  
 QY 1515 ----- 1515  
 Db 12858 CGGCAAGCGAGCGGAGGACCAAGACTCCAGCCCTCTGTGCGGCCACATCTTTC 12917  
 QY 1516 ----- 1516  
 Db 12918 CTCCCCCAGGCTCTTGAACTGAGCGGGGCGCTGCTACTAGGCGGGGTGCTGACCTGGCC 12977  
 QY 1532 OGISePheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAs 1552  
 Db 12978 CGAAGAGCTTCCAGTCCGAATGCGGCAAGTTGTGGGCTGCATGGGGAACCTGCAGGTGGA 13037  
 QY 1552 PSeTArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrVal----- 1568  
 Db 13038 CAGCGGCACTAGACATAGCTGACTTCAATTGCCAACATGGCACCGTGCCTGTATAGG 13097  
 QY 1568 ----- 1568  
 Db 13098 GGGCCGGGGTGGAGCCAGGCTGGGATCCAGTCTGAGAGGAGGGGCTGTGCTGTG 13157  
 QY 1568 ----- 1568  
 Db 13158 AGGATGCCAGAAATGAGAGGTCTCAGAGGCTCTTAAGTGTGCAAGCTTCTCTTT 13217  
 QY 1569 ----- 1569  
 Db 13218 AGAGCCCAAGTCAAGTCACTGAGCTGCTTCTCCATTCAGAGGCTGCGCTGCAAGA 13277  
 QY 1575 YbaenValCysAspSerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAsp 1595  
 Db 13278 AGAACGTGTGTGACAGCAACTTGTCCACATGGGGGCACTTGTGTGAACAGTGGGAGC 13337  
 QY 1595 IapSerCysGluCysProLeuGlyPheGlyGlyCysValAsn----- 1611  
 Db 13338 CGTTCAGCTGGAAGTGCCTTGGGCTTGGGGCAAGAGCTGCGCCAGGATGAGAGG 13397  
 QY 1611 ----- 1611  
 Db 13398 GCGGCTGTAGAGGCAACAGCTGGGTGTCATCAAGTCTGGAGACTGGCAGGGTTGG 13457  
 QY 1611 ----- 1611  
 Db 13458 GCAGGCACTGGGCAAGGCTCTGTGAGCGGGGCTGTGGGTGAAGACGTGTGGCAG 13517  
 QY 1611 ----- 1611  
 Db 13518 AGCCCTGAAGGGCAACAGAGAGCAAAAGGGGCAAGTCAAGAGCAGGCTGGGCAAGC 13577  
 QY 1611 ----- 1611  
 Db 13578 GTGGGGGGGTCCAGGCGAGGTACGCACTTGAAGGCGGGGCTGATGAGGGGAGTGGGC 13637  
 QY 1611 ----- 1611  
 Db 13638 TCTGCTCCCGTGCAGCGGCCACCGCTGAGCATCAAGCCCAAGGCGCTCTGAGGCTCC 13697  
 QY 1612 ----- 1612  
 Db 13698 ACCGTCACAGTGCCTTTCTGCTCTCCAGAAATGGCCAAATTCACAGCACTTCTCTG 13757

QY 1621 GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrLeu 1640  
 Db 13758 GGCAGAGCCTGTGGTGGCTGGGCACTGGCTCTCCCTGCCATTCACCAACCTGTGACTTC 13817  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 13818 AGCCTCATGTTCCGACAGCGCCAGCCAGCGAGTGTCTGTGACAGGCCATCACAGGGGG 13877  
 QY 1661 ArgSerThrIleThrLeu----- 1666  
 Db 13878 CGCAGCACTACACCTCTACAGGTATGATGAAAGGGCGGCTGGCCCTGGCCATTA 13937  
 QY 1666 ----- 1666  
 Db 13938 GGGCCCTGTAGCCTTAGCGGCTGAGACAGAAATGGCTGGGCAAGTCTGGGCAAGGGT 13997  
 QY 1666 ----- 1666  
 Db 13998 GGGGACATATAGAGCGCGCTGATCCGTTGGGAAGTCAATGCTGCCACCTGTGGGCT 14057  
 QY 1666 ----- 1666  
 Db 14058 GAGGAAATATATACGCTTTCCTCTGATGCGCTCAGCTGGGCAAGGGGATGGGTTGAT 14117  
 QY 1666 ----- 1666  
 Db 14118 GCTCCAGAAAGGGAACTGTGAGCTGAGAACAGATTAAGCGCTGTGACCTGT 14177  
 QY 1667 ----- 1667  
 Db 14178 TTTCTTTCCTCTCGGTGTCAAGTACAGAGGGGCAAGTATGCTGAGCGTGGAGGGGCA 14237  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrpHis 1700  
 Db 14238 GGGCTCAGGCGCTCTCTCGGTGTGAGGCAAGCGCGGCGCAATACCGTGTGCTGGCAC 14297  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 14298 CATGCACAGCTGCACATGGGAGCCAGCGGGGGCCCGGCAATGCCATTCGTCTTCGAT 14357  
 QY 1721 TyrGlyGlnGlnArgAlaGlyGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 Db 14358 TATGGGCAAGAGACAGAGGCAACTGGGCGCCCGGTGATGTCTGCACCTGAGC 14417  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
 Db 14418 AACATTAAGTGGCGGAATACCTGGGCGAGCGGAGTGGCCCGTGGCTTTCGGGGC 14477  
 QY 1761 Cys----- 1761  
 Db 14478 TGTTCAGAGTGAAGTGTCTGCGCTGCGCTCCATCCCTCCCAACCACTGACAGCCCC 14537  
 QY 1762 ----- 1762  
 Db 14538 GCTCCACTAGGGCAACACTGCTCTGTCTCTCCAGGGGTGGGGGTAGGCAATACGCC 14597  
 QY 1771 OGILGlyValAsnSerLeuAspProSerHisGlyGlySerIleAsnValGlyGlnGlyCys 1791  
 Db 14598 GGAAGGGGTTTAAACAGCTGGAATCCAGCAATGGGAGAGACATCAAGTGAAGAAAGGCTG 14657  
 QY 1791 AserLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAs 1811  
 Db 14658 TACCTGCTGACCCCTTGTGATCAAAACCGTGTCTGTCAACAGTATATGCAAGCAACGA 14717  
 QY 1811 PTPAspSerTyrSerCysSerCysAsp----- 1820  
 Db 14718 CTGGGACAGCTATTCCTGACGCTGTGA--TCAGAGTATGCTAAGATTCAGGGAACGGGC 14776  
 QY 1820 ----- 1820  
 Db 14777 AGGTATCAGTGTGCTGGGGCAACATGTGCTGTGTGTGATCCTGGGCAATGGGGA 14836  
 QY 1820 ----- 1820

Db	14937	TCACACCACTCTCTGAGTCTTAGTTCCTACACAAACCAATGCGACTGATAGGA	14896b
QY	1820	-----	1820
Db	14897	TGCTATGTGAAAAGAAACAGATGAGGGTTTGATGGAATGGTGCAGGGAGGGTGG	14956b
QY	1821	-----ProGlyTYYTYYGlyAspAsnCyethrAsnVa	1831
Db	14957	GGAATGAGCCTCTCTGGTCTTCTGGTCCAGGGTTACTAATGAGTACACTGATATGT	15016b
QY	1831	1CyAspPLeuAsnProCySgIuhIsgInserValCyethrArgLysProSerAlaProHi	1851
Db	15017	GTGTGACCTGAACCCGCTGTAGACACACAGCTGTGTGTATCCCGAAGCCAGATGCCCA	15076b
QY	1851	sgLYTYTTh-CySgIuCySbProbAntyTleuGlyProTyrcySgIuThr-----	1868
Db	15077	TGGCTATACCTGCGAGTGTCCCCCAATTACTTGGGCCATATCTGTGAGACAGGTAAAC	15136b
QY	1868	-----	1868
Db	15137	AGACCAAGGCAATGTGGCAGACAGTCCCAATGGCTGCTTCTCTGGTGTGCTCTCA	15196b
QY	1868	-----	1868
Db	15197	GAGCCCCGAAAGCCTGGCTGATCCACAGCCAGGGTCAAGAGGGCCACATAGGGCTACCT	15256b
QY	1869	-----ArgIleA	1871
Db	15257	AGGTTAGGTGGAGATGGCAGGGGAGGCTCATGCTCACTGGGTCCCTCTTCGACGGATTG	15316b
QY	1871	bpgInProCySbProArgGlyTThTPrGlyVhiSbProthrcySgIyPrProCyAsnCyAspV	1891
Db	15317	ACCGCCCTTGTCCTCCGTGGCTGGTGGGAGACATCCACATGTGGCCCAATGCACTGATG	15376b
QY	1891	a1SerIySgIyPheAspProAspCySAsnIySthrSergIyGluCyShiSgVlyS----	1909
Db	15377	TCAGCAAAAGGCTTTGACCCAGACTGGACAGCAAGACAGCGCGAGTGCACCTGAAAGTGA	15436b
QY	1909	-----	1909
Db	15437	CAGCCCAAGCAAGCCTCCACTGTGGCCACTTGGGCTCTGTCCAGACTCTCGGGCCCT	15496b
QY	1909	-----	1909
Db	15497	AGAGGGCAGGCTGTAGGAGATGGGCTTCTCTGTAGAGCTAGGGGCTTGTGCTTGTCTCT	15556b
QY	1909	-----	1909
Db	15557	CACAGGACAGGCAAGTTGGCTGCTTACCTTGTTGTGTCTAGAGGGGCAAGGGGACCCCTGGGG	15616b
QY	1909	-----	1909
Db	15617	CAATGTGTAAACCCATGAGCTCTGGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	15676b
QY	1909	-----	1909
Db	15677	TGTGACATCTCGCTGATGGGCTCAACTCCAAAGCATTGTAGTGTATTAATTGTGAATA	15736b
QY	1909	-----	1909
Db	15737	CTGTCTCGAAGGGGACCTGAGGTGCGGACCTCCACAGAGGGGCAAGCTGAGAAAGAC	15796b
QY	1910	-----G	1910
Db	15797	CCATGAGCATCTGGGAAAGAGAGCCTGGCCAGTGAACACCGCTTCTCCCTCTTTCAGG	15856b
QY	1910	-----	1930
Db	15857	AGAAACCACTACCGGCCCCCAGGAGGCCCAACCTGCGCTCTTGTGTGACTGTATCCCAAG	15916b
QY	1930	1ySerIeUSeArArgValCyAspProGluAspGlyGlnCySProCyVlySProGlyValIi	1950

D	b	15917	GCTCCTTGTCCAGAGTCGTGACCCCTGAGAGATGGCCAGTGCTCAGACCAAGCGATGCA	15976
O	y	1950	1eqlvarglncysapaaqcyasapaanProPheAlaIglValThrThraaglYcgg	1970
D	b	15977	TGGGGCGTCAGTGTGAACCGCTGTGACMACCTTTGTGTAGGTTCACCAATGGCTGTGG	16036
O	y	1970	Iu-----	1970
D	b	16037	AAGTGTGGGCTCCTGGGAGATGGGTGGCACCCCTCTTAAGTGTGTAGGCACCTGACC	160986
O	y	1970	-----	1970
D	b	16097	CCAAGAAGCCAGAAAGGGGTGTGTGACGGCTGGGAGGCCCTGGAGGCTTAGAGAG	161586
O	y	1970	-----	1970
D	b	16157	GGCTGGGAGACTGGGAGAGACGACGAGAGGAGAGCTGCTCTGGGTACCATGTGCTCT	162186
O	y	1971	-----ValAantYraapSerCyaproArgAlallegluhalaglyletPTProA	1988
D	b	16217	TCCCCGAGATTAATTATGACGTGCCCCAGAGCGATTAAGCTGGAGATCTGTGGCCCC	162766
O	y	1988	rghrharqPheglYleuProAlaalAlaIProCyafProlysglyserPhe-----	2004
D	b	16277	GTACCCGCTTCGGSCCTGCTGTGCTCCCTGTGCCAAAGCTTCCTTGTGTAGGTGTT	163366
O	y	2004	-----	2004
D	b	16337	GGAGGCCCGATGTGATGTGAGGACATAGGCGCTGTGTGTAGTGGGATGAGGGAGAAA	163966
O	y	2004	-----	2004
D	b	16397	GCTCTGGCTGAAGATCCGGGAGCCCACATTCTGGCTTCAGATCTGGATTCATT	164566
O	y	2004	-----	2004
D	b	16457	CCCATCTTCTGTGATTCACATCCCCCTTCCCATGTGCCAGGTGTGCTTACATCTTG	165186
O	y	2004	-----	2004
D	b	16517	GACTGTTTTGTTTTGTTGTGTGTTTTTTTTTCTGGGACAGAGTGTGTCACTGTACCC	165766
O	y	2004	-----	2004
D	b	16577	CAGGTGTGTGACGATGGCGGTATCTCGCTGCTGCAACTCTGCTCCAGGTCAAG	166386
O	y	2004	-----	2004
D	b	16637	CAATTTTCCGTGAAGCCTGTGGCGGCACACACACTGTGGTAATTTTTTTTTTTTT	166986
O	y	2004	-----	2004
D	b	16697	TTTTTTTTTGTATTTTGTAGAGACAGGGTTTCAACATGTGGCCAGTCTGTGTGAA	167586
O	y	2004	-----	2004
D	b	16757	CTCTTACCTCANTGATCCGCTGCTCGGCTCCCAAAGTGTGGATTAAGGCGTGAAG	168186
O	y	2004	-----	2004
D	b	16817	CAACTGACCCACACACATTTTGGACTTTTTTTTTTTTTTTTGTAGACGAGATCTGTCTC	168766
O	y	2004	-----	2004
D	b	16877	TGTATCAGGCTAACAAATGACGTGGCATATCTGGCTCTGTGCACCTCCACTACAGG	169386
O	y	2004	-----	2004
D	b	16937	TYCAAGGATTTCTCTACTCAGCTTCAGGTCCGAGTGTGNACTACAGGCGTGTGACA	169986
O	y	2004	-----	2004
D	b	16997	TGCCCACTAATTTTTTATATTTTTTTTGTAGACAGGGTTTCAACATGTGGCCAGGA	170566

QY 2004 ----- 2004  
 Db 17057 TGGTCTCATCTCTTGAACCTTGATCAGCCCGCTCGGCTCCCAAGTGGCAGATTA 17116  
 QY 2004 ----- 2004  
 Db 17117 CAGGCGTAGCCACCGTGGCCAGCCACATCTTGAACCTTTAAGAAAAATGCTGA AAAAC 17176  
 QY 2004 ----- 2004  
 Db 17177 CTGCACTGTGCTTGCAATAGGCGCTGTGTAAGATTCAAAATGATCATGCAAGGTTCC 17236  
 QY 2004 ----- 2004  
 Db 17237 AATCGTCACTCAACCAAGTCCTGAGATTACGCCACCAAAAGGCAAGACATCTGGG 17296  
 QY 2004 ----- 2004  
 Db 17297 AGTAGGGGGTGTGGCTGTGACCCAGACACCCAGCCCAAGATGGGGGCTGGCGTGGT 17356  
 QY 2004 ----- 2004  
 Db 17357 CCTGACTGCAATGAGCAGGTTTCAATAGGCAAGGAGTGGGGAGAGCACAAGTGGTGGC 17416  
 QY 2004 ----- 2004  
 Db 17417 TGGAGACAGGACAGATGGGGGCTAAGAGAGTGTATGCTGTTCCTTAGAGCCATG 17476  
 QY 2004 ----- 2004  
 Db 17477 TGTGACATGATCTTGGGAGATTTTGGCCCTGTGGTCTTAATACATGCTTCTCTGTT 17536  
 QY 2004 ----- 2004  
 Db 17537 CACAGAGGCGCTTGTGTTGCTAGTTACTGTCCCGGCGCATGAGGTGGCGCTCCAGCAG 17596  
 QY 2004 ----- 2004  
 Db 17597 AACCTCTCAGCCAGTCTCAGGGGCTTCTGTGTCTCCCTGAGGCGGCGCTTTTGCT 17656  
 QY 2005 ----- 2005  
 Db 17657 TCCCTCCCGCAGGAGCTGTGTGCGCCACTGTGATGAGACAGGGGGTGGCTCCCGCCAA 17716  
 QY 2021 ~~g~~lythralValArgHisCysAspGluHisArgGlyT~~leu~~ProPro 2021  
 Db 17717 ACCTCTTCACTGACAGTCCATCACTTCTCAGAACTGAAGGGCTTCGTAAGTGAACCC 17776  
 QY 2036 ----- 2036  
 Db 17777 CTGATCTCATCTTTTCCCTGTCTTCGTCGTCGAGTCTCACTTGGCCCACTCCCATCTT 17836  
 QY 2036 ----- 2036  
 Db 17837 TGAAGAACGGGGCTTCTGAATTCAGCCGTGTGTCTTGGGCTCCCACTGAGAGGCC 17896  
 QY 2036 ----- 2036  
 Db 17897 GTCTCTCACTCTCGAGGTCCTTTTGTCTCAGAGTTCCGACCTCACTCTGCTCTCCG 17956  
 QY 2037 ----- 2037  
 Db 17957 CTCCTCTCTGTAGGCTGAGGGCTACAGCGAAATGAGTCAAGGCTTGAACCTCAGGGGCGC 18016  
 QY 2052 ~~S~~erGlnGlnLeuAlaLeuLeuArgAsnAlaThrGlnHisThrAlaGlyT~~leu~~ProPheGly 2071  
 Db 18017 TCCGACGAGCTAGCCCTGCTCTCGGCAAGCCCAAGCAGACACAGCTGGCTACTTCGGC 18076  
 QY 2072 ~~S~~erAspValI~~ys~~ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGly~~S~~erThrGln 2091  
 Db 18077 AGCGACGTCAGAGGTGGCTTACAGAGCTGCGCAGCGGCTGGCCCAAGAGCACCCAG 18136

QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGln----- 2106  
 Db 18137 CGGGGCTTGGGCTGTGTGCCACAGAGCGTCACTTCACTGAGGTGGGCTTGAGAGA 18196  
 QY 2106 ----- 2106  
 Db 18197 TGCAGGGCTGGCTGTGTAGATAGGGGTCAATGTAGTGAACCTGTCTATGGCACTGGG 18256  
 QY 2106 ----- 2106  
 Db 18257 GGGCAGAGGGGGCCCTCCATCCCACTACAGAGAGCTCCCTGTGGCAGAGAGGCTCCA 18316  
 QY 2106 ----- 2106  
 Db 18317 TGAACCTGTGACCCCTGAGCCAGCCCTTCCCAACCTTCTCATAGTAACATATGTG 18376  
 QY 2106 ----- 2106  
 Db 18377 TGTGGGGCATGTGTTGACCCAGACAGAGAGGCTGCTGTGGTACATGGGCTGGAGAGA 18436  
 QY 2106 ----- 2106  
 Db 18437 AGCACTATCTGAGACATGTGTGTGTGTCCAGGCAAGGGCTGGAGACTTATGTAGA 18496  
 QY 2106 ----- 2106  
 Db 18497 GAATGAGAGAGAGGCTTAGGGGCAAGACATCACTGACTGAGTGGGCAAGTCCGCTCA 18556  
 QY 2106 ----- 2106  
 Db 18557 GGAGCATTTACAGGGAGGCAACAGTGAACAGAGGAGGGGCTGTGGGGAGCAGGGA 18616  
 QY 2106 ----- 2106  
 Db 18617 CTGGCCGGGCAAGGCGGGGCTCCATGTGTGGGATGTCAAGTGTGGGTTGTAGACA 18676  
 QY 2107 ----- 2107  
 Db 18677 CCCACCGTGAACCTTGCCCAACCCAGAAATCTGCTCGGGTGGGCAAGCGCTCTCTGAGAC 18736  
 QY 2118 ~~F~~AlaAsnLysAspArgHisT~~leu~~LeuGlnGlnThrGlnuGlyP~~leu~~AlaT~~leu~~Leu 2138  
 Db 18737 AGCCAAAGAGGGCACTGGAGCTGATTCACAGACAGAGGGGTGGCAACGCTGCTCT 18796  
 QY 2138 ~~u~~GlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSe 2158  
 Db 18797 CCAGCACTATGAGGCTTACGCAATGCGCTGGGCCCAAGACATGGGGCAACCTACTAG 18856  
 QY 2158 ~~r~~ProPheThrIleValI~~leu~~ProAsnIle----- 2167  
 Db 18857 CCCCTTCACTATGTCAGGCCCAATGATGTAGGCTGTGCTGGTGGGAGGGGTT 18916  
 QY 2167 ----- 2167  
 Db 18917 TGTGAGGAGAGTCCCGCAAGAGCGGCTGTGCGGGTCTGCTGCTCAAGGTCG 18976  
 QY 2168 ----- 2168  
 Db 18977 ATCTGTGACATCCCTTCTTCTTACTCATCTCCGTAAGTCCGCTTGAAGAAAGAACTTGG 19036  
 QY 2180 ~~I~~agIylAlaLysLeuProArgTyrGlnAlaLeuArgIylGlnProProAspLeuGlnT 2200  
 Db 19037 CTGGGGCCAAAGCTCCCGCTACAGAGGCCCTGTGCGGAGACAGCCCGGACCTTGAGA 19096  
 QY 2200 ~~h~~ThrValIleLeuProGlnSerValPheArg----- 2210  
 Db 19097 CAACAGTCATCTCTGAGTCTGTCTTCAAG-AGTCAAGTGTGGCCATGATGAGTTG 19155  
 QY 2210 ----- 2210  
 Db 19156 GAGCTGAGACCCAGTGTCTGTGCAAGTCCACAGAGAGGAGGCCAGCTAAGTGTGAC 19215  
 QY 2221 -----GlnThrProProValValaArgProAlaGlyProGlyI~~leu~~AlaGln 2226

[illegible][illegible]

QY 2557 ----- 2557  
 Db 21435 GTGCTGGGCGATGGGAGAGCAGCATGGAGGCTCATGGCCGAGCTATGGCCTGT 21494  
 QY 2558 ----- ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT 2575  
 Db 21495 CCTATCCCGACG-TCGGGCGTGGCAGCCCTCTCGCCGCTCGCTGGCTGGACGCGCA 21553  
 QY 2575 hTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuPheHisTyrLeuPheA 2595  
 Db 21554 CCGTGGCTGGGCACTGGCTCTCTGTCAAGAGCAAGCCTCTCTTCCACTTCTTTG 21613  
 QY 2595 laThrCysAsnCysIle----- 2600  
 Db 21614 CTAACCTGCAATTGGCATCCAGGTACTGGCCAGCCTGTGGAGAGGAGGACCTGGGCT 21673  
 QY 2601 -----GlnGlyProPheIleP 2606  
 Db 21674 GTGGATGCTGAAATATGACAGACCGTGTGCTGCTTGGCTGGCAGGGCCCTTCATCT 21733  
 QY 2606 heLeuSerTyrValValLeuSerTyrGluValArgLysAlaLeuLysLeuAlaCysSerA 2626  
 Db 21734 TCCTCTCTATGTGGTGTGCTTACAGAGAGGTCCGAAAGCACTCAAGCTTGGCTGGAGCC 21793  
 QY 2626 rGlyAsProSerProAspProAlaLeuThrThrLysSerThrLeuThrSer----- 2642  
 Db 21794 GCAAGCCCGAGCCCTGACCTGTCTGACCAAGTCCACCTGACCTCGGTGAGGAGAGC 21853  
 QY 2642 ----- 2642  
 Db 21854 CAGGGGTCTCAGAGAGCGCGGTAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGG 21913  
 QY 2642 ----- 2642  
 Db 21914 TTGGGCCCACTCCCTTCTCTTTCTCCATCCCTCTCTGAAAGGTGAGAGGAGGAGT 21973  
 QY 2642 ----- 2642  
 Db 21974 AAATGTGTGTGTGGCTGGGAGAGAGATGTTGAAAGGGTGTAGATTGACTTGA 22033  
 QY 2642 ----- 2642  
 Db 22034 GGGAGATGGGCTGAGACAGATATCAAGTAATAAGTCTTGACCTGAGAATCTCTCA 22093  
 QY 2642 ----- 2642  
 Db 22094 TGAATTTAGTAATTTTCTGTCTCTCCAAAGTGTCTCAGAGAGCAATTTCTGTCTCAT 22153  
 QY 2642 ----- 2642  
 Db 22154 CTTTTCAGAGCCATTATATAGTAGTCCAGGATCCCTGGGGGAGAGAGGGGTAGACA 22213  
 QY 2642 ----- 2642  
 Db 22214 GCTGTGGAGAGCTGTGACCACTGCGAGCCTCAATATGGGAGCAGCGAGAGAGAGGAG 22273  
 QY 2642 ----- 2642  
 Db 22274 CCTCTCTTTCAGACCTGAGCATGTGGAGAGTGAACCTTAGGGCAAGTTCCCTCCACCTC 22333  
 QY 2642 ----- 2642  
 Db 22334 CTTCTTCCAGCCTGGGGCGAGCCATCCCATCCCACTTACTGACTCTCTGTCCCTG 22393  
 QY 2643 -----SerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyA 2661  
 Db 22394 CCTAGTCTTCAACACTGCCCCAGCCCTTAGCAGATGGGGCGGTGTACCAAGCCTTAGCGAG 22453  
 QY 2661 sPserAlaGlySerLeuHisSerThrIserArgSerGlyLysSerGlnProSerTyrIleP 2681  
 Db 22454 ACTGGCGGGGCTCTGTGACAGACCAAGTGTGTGGGAGAGGTAGAGCCAGCTACATCC 22513

QY 2681 roPheLeuLeu----- 2684  
 Db 22514 CTTTCTTCTAGAGTGAATCCCGAGATGGAGGGTGGAGAGAGGAGAGGAGGCCAGC 22573  
 QY 2684 ----- 2684  
 Db 22574 CATGTGAGCCAGCGCAGCCAGCTGTGGAGTTGAGAGACACACTGTGGCTGACGT 22633  
 QY 2684 ----- 2684  
 Db 22634 GGGGCGAGCTTGATTGAAGCTGTAAAGGACCACAGCAGAGAAACAGATCCAGGGG 22693  
 QY 2684 ----- 2684  
 Db 22694 AGAGAGAGACTGGGACCTTGGGAGAGGGCGCAGGCTGACCCCTCCAGCATGTCTATC 22753  
 QY 2685 -----ArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLysProG 2703  
 Db 22754 TTCCTAGGAGAGAGTCCGCACTGAACCTGGCCAAAGGCGCCCTGGCTGGGGGATCCAG 22813  
 QY 2703 LysLeuPheLeuGluGluValAspGlnGlnIleAsp----- 2715  
 Db 22814 GCAAGCTTCTCTGAAAGTCAAGACAGCAGCATGC-TGAGACAGAACCTCTGGCCA 22872  
 QY 2715 ----- 2715  
 Db 22873 CCCAGAGAGGAGTGTGGCTGCTTTTACTGAAGGTGGGTGGAGGTGCTGGGCTGGCTGT 22932  
 QY 2716 -----ProAspThrAspSerAspSerAspLeuSerLeuGluA 2728  
 Db 22933 GATCTCTCCCTGGCCTCTAGATCTGACAGACTCCGACAGTGAACCTGTCTTAGAG 22992  
 QY 2728 sPAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlyGluGluGluG 2748  
 Db 22993 ACGACAGAGTGGCTCTTAAGCTCTTACCACTATCAAGACGTAGAGAGAGAGAGAGAG 23052  
 QY 2748 LngLugLugLugLugLalaIlePheProGlyGluGluGlyTyrPAspSerLeuLeuGlyProG 2768  
 Db 23053 AGAGAGAGAGAGAGAGGCGCCTTCCCTGGAGAGCAGGCGTGGATAGCTGTGGGGCCTG 23112  
 QY 2768 LysAlaGluArgLeuProLeuHisSerThrProLys----- 2779  
 Db 23113 GAGCAGAGAGACTGCCCCCTGACAGTACTCCCAAGGTGGGCCAGCACTGGGGCTGTGGC 23172  
 QY 2779 ----- 2779  
 Db 23173 CTTTGGGGCCAGTGGAGAGACAGTGGGCTTGGGGAAGCAGACTGGGGTGTGT 23232  
 QY 2780 -----AspGlyG 2782  
 Db 23233 GGCTCTGTGCTATCTGCCCTGTGTGGGCTCATCTTCTTCCCAACAGATGGGG 23292  
 QY 2782 LysProGlyProGlyLysAlaProThrProGlyLysAspPheGlyThrThrAlaLysGluSer 2802  
 Db 23293 GCCAGGGGCTGGAGAGGCCCCCTGGCCAGAGACTTTGGAGCAACAGCAAAAGAGGTA 23352  
 QY 2802 erGlyAsnGlyAlaProGlyGluArgLeuArgGluAsnGlyAspAlaLeuSerArgGluG 2822  
 Db 23353 GTGGCAACGGGGCCCTGAGAGAGGCGCTGCGGAGAAATGAGATGCCCTGTCTGAGAGG 23412  
 QY 2822 LysLeuGlyProLeuProGlyLysSerAlaGlnProHisLys----- 2836  
 Db 23413 GGTCCCTAGGCCCCCTTCAAGGCTCTTCTGCCAGCTTCAAAAGGTAGTGGGACACC 23472  
 QY 2836 ----- 2836  
 Db 23473 CCAAGTCCGAGCTCCCTAGTCAAGCAGCCTCATACTCATCTTCTGTGGCCGAGCC 23532  
 QY 2837 -----GlyLysLeuLysLysCysLeuProThrIleSerG 2849  
 Db 23533 TCACAGCCCGGCGCCGAGCCACAGGAGTCTTAAAGAGAGTGTCTGCCACCATCAGCG 23592  
 QY 2849 LysLysSerSerLeuLeuArgLeuProLeuGluGlnCysThrGlySerSerArgGlySers 2869





Db 321 GGGTCCAGGGGACGAGGCTCTTTCGGGGGCGTGGCCCCCATGGGCTGCTGTCTCATCC 380  
Qy 61 SerAlaSerAsnLeuThrLeuThrLeuThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
Db 381 TCAGGCTCCAGACTCTGGCTCTTACACAGCCGCTGCAGGGATGCGGGCATGAGCTGACT 440  
Qy 81 GlyHisLeuValProHisAspGlyLeuArgValThrCysProGluSerGluAlaHis 100  
Db 441 GGCCACTGGATCCCAACCAACGATGAGGCTGGAGGGTTGGTGCACAATCCAGAGGCCAT 500  
Qy 101 IleProLeuProProAlaProGluGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
Db 501 ATTCCTCCACACAGACTCTGAAAGGCTGCCCTGAGCTGTGCTCTCTGGGCATTGGA 560  
Qy 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGluHisProCysLeuLeuVal 140  
Db 561 GGCCACCTTCTCCCAACAGGGCAAGCTCACACTGCCGAGAGAGACCCGATTAAGGCT 620  
Qy 141 ProArgLeuArgCysGluSerCysIleuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 621 CCAAGGCTCAAGATGCAAGTCTGCAAGCTGCACAGGCCCGGGGCTCAGGGCAAGGGAA 680  
Qy 161 ArgSerProGluGluSerLeuGlyGlyArgArgLeuAspAsnValAsnThrAlaProGln 180  
Db 681 AGGTCAACGAAAGATCCCTGGGTGGCGTCCGAAAGAAATGTAAATCAAGCCCCCAG 740  
Qy 181 PheGlnProProSerTrpGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
Db 741 TTCACAGCCCCCAAGCTACAGAGCCACAGTCCGAGAAACAGCAGCAGCAGCAGCCCTGTT 800  
Qy 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluThrMet 220  
Db 801 GCATCCCTGAGGGCCATTCACCCGAGAGGGTGAAGGACAGTGCAGTGAATGACACAG 860  
Qy 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
Db 861 GATGCCCTCTTGATAGCCGCTCCAAACAGTTCTTCTCCCTGACCCAGTCACTGTGTGA 920  
Qy 241 ValThrThrAlaGluGluLeuAspArgGluThrIleValPheArgValThr 260  
Db 921 GTAAACCAACGAGGAGCTGATCGTGAACCAAGAGCACCAAGTTCAGGGGTACAG 980  
Qy 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
Db 981 GGCAAGAACACAGGACATGCCGACAGAGTCCCTGGCTACACTCACTCTTGTGTTACT 1040  
Qy 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluThrIleGlyLeuSerLeuArgGlu 300  
Db 1041 GACACCAATGACATGACCTCTGTTCAGACAGCAGAGTACAGAGAGCTCCAGGAG 1100  
Qy 301 AsnLeuGluValGlyTrpGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 1101 AACTCGAGGTGGCTATAGAGTGTCACTGTCAAGGCAACGAGATGTATCCCTCC 1160  
Qy 321 AsnAlaAsnIleLeuThrArgLeuLeuGlySerGlyIleSerProSerGluValPhe 340  
Db 1161 AATGCAATATCTGATCCGCTGCTGAGGGGTCTGGGGGCAAGCCCTCTGAATCTTT 1220  
Qy 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
Db 1221 GAGATGAGCCCTCGCTCGGGGTGATCCCAACCGTGGCTCTGTGATCGGAAAGGTG 1280  
Qy 361 GluSerTrpGluLeuThrValGluAlaSerAspGlnIleArgAspProGlyProArgSer 380  
Db 1281 GATCTCTACCACTGACGGTGAAGGCAAGTGAACAGGGTCGGGACCCGGGTCTCTCGAGT 1340  
Qy 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
Db 1341 ACCAAGCCGCTGTTTCTTCTGTGAGAGATGACAAATGATTAATGCCCCCACTTATGT 1400  
Qy 401 GlnIleArgTrpValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
Db 1401 GAGAAAGCGCTATGTGTGTCAGGTGAGGAGGATGTGACTCCAGGGGCCCACTACTCCA 1460

Qy 421 ValThrAlaSerAspArgAspIleGlySerAsnAlaValIleHisTrpSerIleMetSer 440  
Db 1461 GTCAAGCCTCGATCGACAGACAGGGAGCAATGCCGTGGTGCATATATGATCATAGAT 1520  
Qy 441 GlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyValLeuAspValIleSer 460  
Db 1521 GGCAATGCTCGGGGACAGATTTATCTGATGCCAGACTGGAGCTCTGTGATGTGTGAGC 1580  
Qy 461 ProLeuAspTrpGluThrThrArgGluThrLeuArgValArgAlaGlnAspGlyGly 480  
Db 1581 CCTCTTGACTATGAGACACACAGAGTACACCTTACGGGTGCGACACAGATGTGGC 1640  
Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1641 CGTCCCCACTCTTAATGTCTCTGGCTGGGACATGACAGTCTCGATATCAACAGC 1700  
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
Db 1701 AATGCCCATCTTCTGTGACAGACCCCTTCCAGGCTACTGTCTTGAGAGTGTCCCTTA 1760  
Qy 521 GlyTrpLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1761 GGCTACCTGGTCTTCATGTCCAGGCTATCAAGCTGATGCTGGTGAATGCCCCGCTG 1820  
Qy 541 GlnTrpArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1821 GAATACCGCTTGTGGGTGGGATGACATGATCCCTTCCATCATCAATGAGCAGAGC 1880  
Qy 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTrpSerPheGlyVal 580  
Db 1881 TGGATCTGTGTGGCTGTGAACTGACCGGAGAGAAATGTATTTCTACGCTTGGGGTA 1940  
Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
Db 1941 GAAGTCAAGACATGGACATCCAGCACTCACTGCTGGCCAGTGTCAAGCTGACTTC 2000  
Qy 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTrpValArgLeuAsn 620  
Db 2001 CTGGATGTCACACGACAAATCAATCCAACTTTACCAACAGAGTACACAGTGGCTCAT 2060  
Qy 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
Db 2061 GAGATGACGCTGTGGGACCAAGGTGTGACGGTGTGACGTGTGACCGGTATGCTCAT 2120  
Qy 641 SerValIleThrTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
Db 2121 AGTGCATCACTTACAGATCACAGTGGCAATCTGAAACCGCTTCCATCACACAGC 2180  
Qy 661 GlnSerGlyGlyIleLeuValSerLeuAlaLeuProLeuAspTrpIleLeuGluValGln 680  
Db 2181 CAAAGTGGTGGGTGTGTATCCCTTGCCCTGCACTGAGCTCAAACTTGAAGGGCAG 2240  
Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
Db 2241 TATGTGTGGCTGTACCGCTTCGATGGACCTGGGAGGACACAGGACCAATGTGTGTG 2300  
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTrpThrValAsn 720  
Db 2301 AATGTCAACGACGCAACACCACTGCTCTTCTTCAAGCTCCCATATACATGTAAT 2360  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
Db 2361 GTTATATGAGACCGGCGCGGACAGCAACGGTGTGTGATCAACGACCAACGATGAGGAC 2420  
Qy 741 ThrGlyGluAsnAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2421 AAGGTGAAGATGCCGATCACTTCTCATGAGAGCAACCAATCCCAATTCGGCATTC 2480  
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTrpGluAspGlnValSer 780  
Db 2481 GATGACAGCACAGGGGCTGTACCAACCAAGCTGAGCTGATGAAAGCAATGTGTCT 2540

QY 781 TTTThleuAlaIleThraAlaArgAspAsnGlyIleProGlnIlySerAspThrThrTy 800  
DB 2541 TAAACCTGGCCATTACTGCTCGGGAACAATGGCAATCCCGAAGATCCGACCACTTAC 2600  
QY 801 LeuGlnIleIleuValAsnAspValAsnAspAsnAlaProGlnPheIleuArgAspSerTy 820  
DB 2601 CTGGAGATCCCTGGTGAACGACGTGAATGACAAATGCCCTCAGATTCCTGGCAGACTCTAC 2660  
QY 821 GlnGlySerValTyThrIleuAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
DB 2661 CAGGGAGTGTCTATGAGATGAGCCACCTTCACTAGCGTCTTCAAGATCTAGCCACT 2720  
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyIleAspAspGly 860  
DB 2721 GATCGATTTCTGACTTAATGCAAGGCTTCTTCAACCTTCAAGAGGCGACATGGA 2780  
QY 861 AspGlyAspPheIleValGlySerThrSerGlyIleValArgThrIleuArgIleuAsp 880  
DB 2781 GACGGTGACTTATTTGTGTGATCAGTCAGGCACTGTGGAACGCTACCGAGGCTGGAT 2840  
QY 881 ArgGlyAsnValAlaGlnTyValIleuArgAlaTyAlaValAspIlySerProPro 900  
DB 2841 CAGGAACCTGGGCCCGCATGATGTTGCGGCAATGCAAGTGAACAAGGAGATGCCCA 2900  
QY 901 AlaArgThrProMetGlyValThrValThrValIleuAspValAsnAspAsnProProVal 920  
DB 2901 GCCCGCACACTTATGGAAGTGAAGTCACTGTGTGGAATGTAATGCAATCCCTGTGC 2960  
QY 921 PheGlnGlyAspGlyIleuPheAspValPheValGlyIleuAsnSerProIleGlyLeuAlaVal 940  
DB 2961 TTTGACACAGATGATGTTGATGTTGTGTGGAAGAAACAGCCCATTTGGGCTAGCCGTG 3020  
QY 941 AlaArgValThrAlaThrAspProAspGlyIleThrAsnAlaGlnIleMetTyArgIle 960  
DB 3021 GCCCGGCTCACACCACTGACCCCGATGAAGGCAACAATCCCAATTAATATGACGATT 3080  
QY 961 ValGlyIleValAsnIleProGlyValPheGlnIleuAspIlePheSerGlyIleuThrAla 980  
DB 3081 GTGGAGGGCAACCTCCGAGGTCTTTCAGCTGGAACATCTTCCGGGGAGCTGACAGCC 3140  
QY 981 LeuValAspIleuAspTyArgIleuAspArgProGlyIleValIleuValIleGlnAlaThrSer 1000  
DB 3141 CTGTGACTTAACATACGAGACCGGCTGAGTACGCTGTGATCCAGGCAAGCACTCA 3200  
QY 1001 AlaProIleuValSerArgAlaThrValHisValArgIleuAspArgAsnAspAsnPro 1020  
DB 3201 GCTCTCTGTGAGCGGGCTACAGTCCAGCTCCGCTCTTGAACGCAATGCAACCA 3260  
QY 1021 ProValIleuGlyAsnPheGlnIleuPheAsnAsnTyValThrAsnArgSerSerSer 1040  
DB 3261 CCAAGTCTGGGCAACTTGGATCTTTTCAACAACATATGTCAACAATGCTCAACAGC 3320  
QY 1041 PheProGlyIleValIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu 1060  
DB 3321 TTCCTGGGGGTCCCATGGCCGAGTACCTGCCCATGACCCTAATATCTCAGATATCTG 3380  
QY 1061 ThrTySerPheGlnIleuArgIleuAsnGlyLeuSerIleuValIleuAsnAlaSerThrGly 1080  
DB 3381 ACTTACAGCTTTAGGGGGAATGAACCTCAGCTGTCTGTCTCAATGCTCAAGG 3440  
QY 1081 GlnIleuIleuSerArgAlaIleuAspAsnAsnArgProIleuGlnAlaIleMetSerVal 1100  
DB 3441 GACCTGAAGCTTAAGCCGCGCATGACACACACCGGCTCTGGAGGCAATCATGAGCGTG 3500  
QY 1101 LeuValSer----- 1103  
DB 3501 CTGGTGTCAAGTGAAGAGGGCCAGGTGGCGCTGGGGTGGAGTACTCGCGGGAGT 3559  
QY 1103 ----- 1103  
DB 3560 GTCTGGGCAAGCACTGAGGTAAGGGTGCATCCAGGAAGCAAGTACAGATCCACTCCCT 3619  
QY 1103 ----- 1103

DB 3620 GCCCAGTGTCTGGCAACAGAGCAGAGGGGCAAGACCAGCTTGGGAAGAAGCCCAAGAA 3679  
QY 1103 ----- 1103  
DB 3680 CCTGGCGGCTGGTGCAGGCCCCCACTTCTGTGGCTGGGCTGGGCTGCTGTGGTTGC 3739  
QY 1103 ----- 1103  
DB 3740 TGCCTCCACTGTCTACAGCCTGTGACGTGTGCGGGAAAGTGTGTAGAGCTGCCCCGC 3799  
QY 1103 ----- 1103  
DB 3800 TATGCTGCAGAAATTGTGAAAAAAGGCGGTGAGAGCTTGGGGGCAAGAGAGCTGG 3859  
QY 1103 ----- 1103  
DB 3860 GGAAGGCGAAGAGAGCTGTGGCGCAGAGAGAGAGGCGAGGCCCTTCAACCC 3919  
QY 1103 ----- 1103  
DB 3920 CATCAAGATATTCTGCACTCTCAACAGCCTTTGGCAGAGCTGGGGAACACGGAGAGAC 3979  
QY 1103 ----- 1103  
DB 3980 AGGCTGCGCCACCAAGGGGTCAAGAACCTGTAGTCTCTTGTGGCCCTTTTGTCTCC 4039  
QY 1103 ----- 1103  
DB 4040 TGGCTAGCTGCAAGATCGGCTGACAGGAACTTTGTAGCGCATTTTTCCTTCACAG 4099  
QY 1103 ----- 1103  
DB 4100 CAGAAAGCCTGGGCTGTGCTCCCTTCAAGACTCCACAGAAAGTAGAGCCAGAGTTC 4159  
QY 1103 ----- 1103  
DB 4160 ATTGTCTGTCCAAACAGACCCCACTTTCAGAGTCCGTGTTCGGGTTCGCTTGGAGAG 4219  
QY 1103 ----- 1103  
DB 4220 GGGCTGCAATGAAGCTGGGGGGCAGGCTGAGGCTGTGAGGCCCTTGGCCACTCATG 4279  
QY 1103 ----- 1103  
DB 4280 GGTGTCTTAGGTTGCAAGGCTTCTTCTGTAGCTGTGACCAAGCTCCAG 4339  
QY 1103 ----- 1103  
DB 4340 GAATGAAGTGAAGGACAGACAGGCAATGACTGTGACTTTGAGGGCAGAGACCC 4399  
QY 1103 ----- 1103  
DB 4400 CAAAGCACTTATCTGAGAGAGAAACAGAGTCCAGGGCGTGGAGGCTTGGGAAACC 4459  
QY 1103 ----- 1103  
DB 4460 AGGCTGGGCAATTCCTGTGCTTGGCAAGCACTCAGCTTACAGACTCCCTGAGCAGAG 4519  
QY 1103 ----- 1103  
DB 4520 GGTGGGGAGAGAGTTGAGAGGGGTTGTGACAGTATCTGTCTTTTATACAGCTG 4579  
QY 1103 ----- 1103  
DB 4580 TGGGAGCCAGAGTGAAGCAGGGAAGAACCAACCAAGTTGTGATGCTTTCC 4639  
QY 1103 ----- 1103  
DB 4640 AAGCTGGGAACACTGAGACCCTCATGAGGGAGTGTGTAGTCAACCCAGGGGTGG 4699  
QY 1103 ----- 1103

Db 4700 ATCAGGTGCATCCCTCTAGGCAAGCTTCAATCATCAGGGAAGGAGAAAGAGA 4759  
QY 1103 ----- 1103  
Db 4760 GGCCTGACAGGCCCTTACTCCACCCAGGAGATGTGCTGAAGGGAGAGAGGGTATG 4819  
QY 1103 ----- 1103  
Db 4820 CCAGAGTCAGGGCGTGGCTCTACTCAGGGTCTTGTGTGACAGATGCTCCCTGCCCCAC 4879  
QY 1103 ----- 1103  
Db 4880 CTTTCAAGTACCAACAGAACCAAGAGAGCTTTTCAAGGACTGTGCTGAGTT 4939  
QY 1103 ----- 1103  
Db 4940 CTGAGCTCTGAGACTTCAATTTCCCTCTCAGAGAGCGGAAGAGGAGGCTGAGCTCC 4999  
QY 1103 ----- 1103  
Db 5000 TGTGTCTCACTGTGCTGGGCACTTGACAGAGTGGAGAGAGTGGAGAGAGGAGG 5059  
QY 1103 ----- 1103  
Db 5060 TCTCCCTCTGTCTCCACACCCAGCTGCTTTGTGTAGCGTGGCAGGCACTGGCAGA 5119  
QY 1103 ----- 1103  
Db 5120 GAAGGCCCCCATTAAGCTTGTGTCTCCACAGGGGAGAGCCCTAGACTTGTGTTCC 5179  
QY 1103 ----- 1103  
Db 5180 TAGTCTAGAGAGCTCGGGGACATCCAGAGAGGCTCTCTGCGCCGAGAGCTTGG 5239  
QY 1103 ----- 1103  
Db 5240 CCCCACTCTCTCCCTTCGGGCTCTGAGAGGAGTACCAAGGCAACCCCAAGTT 5299  
QY 1103 ----- 1103  
Db 5300 CAGGAGCGCGGGGTGAGGGGAGAGTACTCCACTGCCCCCTTTGTTTCTCTCTGT 5359  
QY 1103 ----- 1103  
Db 5360 TCTTGTCTCTTCTTCTTCCGCGCTGCGCCATGCGAGCCCTGACAGACTGCTCTC 5419  
QY 1103 ----- 1103  
Db 5420 TCCGGTGCCTCTCCCTGTCTGTGGGCTCTCCGCTCTCCACTTGGCCATTGCTGCT 5479  
QY 1103 ----- 1103  
Db 5480 CCTCTCTGAGGCTTGTCTTCCCTGTGTTGTGTCCTGTTGTGCTCACTCTCTTC 5539  
QY 1103 ----- 1103  
Db 5540 CCTATGCTGAGCTGCCAGCTGGGCGAATGCTCCCTGAGGCTGAATCTCTCTC 5599  
QY 1103 ----- 1103  
Db 5600 GCTTCTTCTCTCTGTCTGCTCAGACACAGCAGAGCTCTGTCCATGATCTGGCAGCTCC 5659  
QY 1103 ----- 1103  
Db 5660 AGAACAGAGAGCCTCTTGTGCTCAGTCTCGCTGACATCTGCAAGGAGAGAGAG 5719  
QY 1103 ----- 1103  
Db 5720 GCTGGAGGCTGACCGGAGCTTGACATTTTCCAGTGTGCTGCTAAGCTTGGAGAC 5779  
QY 1103 ----- 1103  
Db 5780 AGGTGGCATTTCTGTGGGTGAGAGGCAATCCAGCTGTCTTTCTGTCCAGAGAGACA 5839

QY 1103 ----- 1103  
Db 5840 GCTCAGGAAGTACTCCCAACCCCTGTTCCTCATCTGCTTTGCACTGAGGGGTGAC 5899  
QY 1103 ----- 1103  
Db 5900 TAGTCAAGCAGGGAGGATGATATAGAAAGAGAGACAGTGGGTTGACTTGGGGA 5959  
QY 1103 ----- 1103  
Db 5960 TGGGGACACACAGCAGAGGGGCGAGTGGCCACAGATGTGTGACCATGAGGACAA 6019  
QY 1103 ----- 1103  
Db 6020 GAGACATGCTGGAGAGGGGCCAAGCGGGGCTTAAGCATGCGCGTGGGGGACACAG 6079  
QY 1103 ----- 1103  
Db 6080 GAGAGTGTGTTTCCAGGCTGAGGGGTGCAAAAGTTCTCCGCTCATGCGCTTGG 6139  
QY 1103 ----- 1103  
Db 6140 TTGCCCCCTCTCTCTTAATCAGTTATGAGCACTGTGTGTGAGTCTTGGGCTGCTT 6199  
QY 1103 ----- 1103  
Db 6200 CTGCTGAGGGAATATCCGCTCAAGAAATCATTAGCGATATATCAGGTATCAAGAG 6259  
QY 1103 ----- 1103  
Db 6260 CTGTGTCACTGAGCATCTGTGTGCGGAGTCTGAAAAGAGAGAAAGTAGAGCC 6319  
QY 1103 ----- 1103  
Db 6320 TGAATCATTAATCTGGCAGGCTTCTGTGACAGTGAAGCTTAGGAGCTTAAAGACT 6379  
QY 1103 ----- 1103  
Db 6380 CAGAGACTGTGGGTGTGTGAGTGGAGAAACAGCTGTGACAGCTTTGAGAC 6439  
QY 1103 ----- 1103  
Db 6440 GAGACAGAGAAACAGATGTGTATGACAGAGCTTGTGTGAGAGTATGTAG 6499  
QY 1103 ----- 1103  
Db 6500 AATTAAGATGATGAGGCTGAGGGTATGTTGGCAATCGGCTGACATGTGATGAT 6559  
QY 1103 ----- 1103  
Db 6560 GGGGTCTGCTTTCAGGCTTCCGGCTTACAGTGAACCGCTGTATCAGTACGACAC 6619  
QY 1103 ----- 1103  
Db 6620 GCAATAGTATCAACACAGATGCAAAATGTTCATGTCCCTTGTCTTCCACCCC 6679  
QY 1103 ----- 1103  
Db 6680 ACCCAGAAACATGGGCTCTGCCCCGTACTGCTAGGTAAGCATTTGCTTTTATATC 6739  
QY 1103 ----- 1103  
Db 6740 TAGTTGGCTCCCTGTATATGTGTTTGTGTTTGGAGACAGATCTCAGCTG 6799  
QY 1103 ----- 1103  
Db 6800 TCACCAAGGCTGAGTACAAATGGCGCATCTCGGCTCACTGCAACTCTGCTCTGGGT 6859  
QY 1103 ----- 1103  
Db 6860 TCAAGCAATTTCTCTGCTCAGGCTCCTGATATCTGGAGCTCTGATGATGATTT 6919

QY 1103 ----- 1103  
Db 6920 ACAGGACCCACACACACGCCAACTAATTTTGTATTTTGTAGAGAGGGGTTTCAC 6979  
QY 1103 ----- 1103  
Db 6980 CATGTGTGCACTGTATCTCGAATCTCTGACCTCTGTATCCACCCGCTCAGCCTCCCA 7039  
QY 1103 ----- 1103  
Db 7040 AAGTTTGGGACCTACAGGCGGTGAGCCACTGCGCGCGCGCGCTGTATGTTTAAAGTCC 7099  
QY 1103 ----- 1103  
Db 7100 GTTATTTCTTTGTGAGACTGAAATAGATTGTATCCACATCTGAGTTTTCGAACCC 7159  
QY 1103 ----- 1103  
Db 7160 TACTATGTGTAAAGCATTTGTAAATPACAAATATGAAATATGCCAGTAGCTCATGCTGT 7219  
QY 1103 ----- 1103  
Db 7220 AATCCAGTGTCTAGAGAGCTGAGATGAGGATCACTTGAGGCCAGAAATTTAGACCA 7279  
QY 1103 ----- 1103  
Db 7280 GCCTGGAACAACAGAGAGACCCCATCTCTACACACAAACAAATTTAGCAGGATAGT 7339  
QY 1103 ----- 1103  
Db 7340 TGTGAGCACCTGTGTCTCAGCTATTGTGGAGGTGAGGTGAGAGATGCTTGAGCCCA 7399  
QY 1103 ----- 1103  
Db 7400 GGAGTTGAGGCTACAGTAGCTATGATGTGCTCCTCCAGTGTAGGCGACAGAG 7459  
QY 1103 ----- 1103  
Db 7460 AGTCTGTCTCTTAAAAAAG 7519  
QY 1103 ----- 1103  
Db 7520 CCTCTTCTCCAGGCGCCAGCTTCAATTTTAAAGTAAATTAATTCATTTCCAGTGTGA 7579  
QY 1103 ----- 1103  
Db 7580 TGTGATTCAGGTGTGTATGTATTTCTGGGCTTAGGGTGTGTGGTGAATGAGACACC 7639  
QY 1103 ----- 1103  
Db 7640 AGCAAGTGTGTTTAGTGTAGCTGTATTAATGAGTGTGTGCTGGGGCATAGCA 7699  
QY 1103 ----- 1103  
Db 7700 GGCAATTTGGGTGTAAATGTGTAGCCAGAGTGGGTATGATGAGTAGAGCCCTAGSACCA 7759  
QY 1103 ----- 1103  
Db 7760 GAGAGAGACTGACTCAGTGGGGCTCTGGGGTCTCAGAGTGAAGAGCTGCCAGGCCCT 7819  
QY 1103 ----- 1103  
Db 7820 CATCTCTGCTCCGGTCCCGACCCGACCTCACTCCTTCACAGATGGGGCAGTGGG 7879  
QY 1103 ----- 1103  
Db 7880 TGAATGATGCACTGTAAAGTGTACGTGCTCAGCTGATTTCTGTGGGTGGGGCCCCAGGCT 7939  
QY 1103 ----- 1103  
Db 7940 ACCCAGAGAGCCAGAGAGTGGGGGAGAGAGATGTTGTGTGATGTTCATGGG 7999  
QY 1103 ----- 1103

Db 8000 GATACAGGTCTGGCGGCCCAAGCAGCATGGGGCTCAGTGGAGGGGGACTAGGTGACC 8059  
QY 1103 ----- 1103  
Db 8060 CACTGACCCAGTTTACCACTCCCCCCCCCACTCCCTGTGTGTGGGGAG 8119  
QY 1103 ----- 1103  
Db 8120 TGGGAAACAGACCATGTGTGCGATTATGTGTGTGTGAGGTGTAGAGCTGAGGGGCA 8179  
QY 1103 ----- 1103  
Db 8180 GGGGATGATTTTGTGGGTCTCAGGATGCTGTGGGCTGTGAGAAAGTATGAAAAAGA 8239  
QY 1103 ----- 1103  
Db 8240 CAATAGGGTTCCCGCAGCAGCAGACAGACTGTGGGGAAAGGAGCAGAGCGGTGCG 8299  
QY 1103 ----- 1103  
Db 8300 TTGGGAGGAAAGGTGCAAAACATTAAAGGTGTGCTGTGATCAGCAGACAGAGTT 8359  
QY 1103 ----- 1103  
Db 8360 TTGGAAAGATGCAGAAAGAACTGCAAAATCTCATCCCTGAGCCCTGAGAGATGTTG 8419  
QY 1103 ----- 1103  
Db 8420 GCCACAGTGCATACCTGTGCTACGACCTTATGCAATTGAGAAAGAGACCTTTCT 8479  
QY 1103 ----- 1103  
Db 8480 TGTGTGTGTGAGGAATATGAGGCTGAAATGAGCCCAAGCTGGGCTCTGACTGTG 8539  
QY 1104 ----- 1119  
Db 8540 TCCCTCTCAGAGAGGGGTACAGAGCTGACCGCCAGTGTGCTGTGTGACATC 8599  
QY 1120 ----- 1139  
Db 8600 ATCACCAGTAGAGTCTACCCACAGATCAGCTGCGCTGTGAGAGACATGTACCCAG 8659  
QY 1140 ----- 1159  
Db 8660 ATGTherLeuSerProLeuLeuGlyLeuPheIleGlnAlaAlaThreuAlaThr 8719  
QY 1160 ----- 1179  
Db 8720 PropProAbpHLeuValValPheAbnValGlnArgAbpThrAspAlaProGlyGlyHis 8779  
QY 1180 ----- 1199  
Db 8780 ATCTCAACGTGAGGCTGTGAGTGGCCAGCGCCAGCGCGCGCGCGCGCGCGCTTC 8839  
QY 1200 ----- 1219  
Db 1200 LeuProSerGluAspLeuGlnGlnArgLeuThyLeuAbnArgSerLeuThraIle 1239  
QY 1220 ----- 1239  
Db 1240 SerAlaGlnArgValLeuProPheAbpAbnIleCysLeuArgGluProCysGluAsn 1259  
QY 1260 ----- 1279  
Db 1260 TyrMetArgCysValSerValLeuArgPheAbpSerSerAlaProPheIleAspSer 1299  
QY 1280 ----- 1299  
Db 1280 PheThrGlyAspGlyTrpCysGluThrGluValAspLeuCysTrpSerArgProCysGlyPro 1299







Db 13458 GCAGGCACTGGGACAGGGCTCTGCTGAGCGGGCTGNGGGTGGAAAAACGCTGCTGGGCGAG 13517  
 QY 1611 ----- 1611  
 Db 13518 AGCCCTGAAAGGGCAACACAGAGACAAAGGGGCCAGTCAAGAGGACGCTGGGCAACAGC 13577  
 QY 1611 ----- 1611  
 Db 13578 GTGGGGCGGTCCAGGGGAGTACGCACTTTGAGAGGGCGGGCTGANTGAGGGAGTGGGC 13637  
 QY 1611 ----- 1611  
 Db 13638 TCTGCTCCCGTGCAGCGCGCACCGCTGAGCATCAAGCCCGAGGCGCTCTGAGGCTCC 13697  
 QY 1612 -----gluMetAlaasnProgluNhiAspheleu 1620  
 Db 13698 ACCCGTCACAGTCCTGCTTTTCTGCTCCCGAAGAAATGGCCAAATCCACAGACCTTCTCG 13757  
 QY 1621 GlysSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrleu 1640  
 Db 13758 GGCACAGACCTGTGGCTGTGACATGGCCCTCTGCTGCCCATCTCCCAACCTGTATACCTC 13817  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 13818 AGCCTCATGTTCCCGACCGCGCACGCGTGTCTGCTGACAGGCCATCAACAGGGGG 13877  
 QY 1661 ArgSerThrIleThrleu----- 1666  
 Db 13878 CCGACACACATCAACCTTACAGGTGATGCAATGAAAGGGCGGCTGGCCCTGGCCATA 13937  
 QY 1666 ----- 1666  
 Db 13938 GGGCCCTGTAGCTCTAGAGCGGCTGACAGAAAGTGGCTGGGAGGTCTGGGCGAGGGGT 13997  
 QY 1666 ----- 1666  
 Db 13998 GGGGACATATAGAGCGCGCTGATCCGTTGGAGAGTCAATGCTGCACCTGTGGGCT 14057  
 QY 1666 ----- 1666  
 Db 14058 GAGGGAATTAATAGCTTCTCTCTATAGGCTCAAGCTGGGACAGGGATGGGTTGAT 14117  
 QY 1666 ----- 1666  
 Db 14118 GGTCCAGAAAGGGAATGTGAGCTGAGAGACATGATTAAAGCGGCTGCTGACCTGT 14177  
 QY 1667 -----GlnleuArgGluGlyNhiValMetLeuSerValGluGlyThr 1680  
 Db 14178 TTTCTTCTCTCTGTGTTGCTAGCTACGAGAGGGCCACGTTGATCTGAGCGGTGAGGGCA 14237  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAlaAsnAspGlyAspTrpHis 1700  
 Db 14238 GGGCTTCAAGGCTCTCTCTCTGCTGAGAGCGCGGCGGCAATGACCGTATCTGGCAC 14237  
 QY 1701 HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyNhiAlaIleLeuSerPheAsp 1720  
 Db 14238 CATGCAACGCTGGCACTGGGAGCCAGCGGGGGGCCCGGCAATGCACTTCTGTCTTCAT 14357  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyValLeuGlyProArgLeuNhiGlyLeuNhiAlaLeuSer 1740  
 Db 14358 TATGGGCGAGCAAGAGGAGGCAACCTGGGCGGCGGCTGCATGCTGTGACCTTGAGC 14417  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 Db 14418 AACATTAACGTGGGGGAAATACCTGGGCGAGCGCGGTGTGGCCCTGTGGGCGC 14477  
 QY 1761 Cys----- 1761  
 Db 14478 TGTTCGAGGTGAGTCTGCTGCGCTGCTCCATCCCTCCCAACCACTGACAGCCC 14537  
 QY 1762 -----Leu-GlnGlyValArgValSerAspThrPr 1771  
 Db 14538 GCTCACTGAGGGCAACACTGCTCTGCTCCCTCCAGGGGTGTGGGAGGATACGCC 14597

QY 1771 OGluGlyValaAsnSerLeuAspProSerHisGlyGlySerIleAsnValGluGlnGlyCys 1791  
 Db 14598 GAAAGGGTTAAAGCTTCGATTCACGATGGAGAGATCAACGTGAGCAAGGCTG 14657  
 QY 1791 sSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAs 1811  
 Db 14658 TAGCTGCTGACCTTGTGTGACTCAACCGGTGTCTGTCTAAAGCTATTGCAAGCAAGCA 14717  
 QY 1811 PTrpAspSerTyrSerCysSerCysAsp----- 1820  
 Db 14718 CTGGACAGCTATTCTCGCACTGTGA-TCCAGGTAATGAAGATCCAGGGCAACGGGC 14776  
 QY 1820 ----- 1820  
 Db 14777 AGTTTACAGTGCCTGGGGCAACATGCTGCTCTGCTGTGTGACTTGGGCAATGGGCA 14836  
 QY 1820 ----- 1820  
 Db 14837 TCACACACACTCTCTGAGTCTTAGTGTGCTTACACAAAGCCATAGCCAGCTATAGGA 14896  
 QY 1820 ----- 1820  
 Db 14897 TGTATGTGAAGAAACAATGAAGGTTGATGAAGATGTGCCAGGGAGGTGGG 14956  
 QY 1821 -----ProGlyTyrTyrGlyValAsnAsnCysThrAsnVal 1831  
 Db 14957 GGAATGAGCCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15016  
 QY 1831 1CysAspLeuAsnProCysGluNhiGlnSerValCysThrArgIleProSerAlaProNhi 1851  
 Db 15017 GTGTGACCTGAACCGGTGTGAGCACAGTCTGTGTATACCGGCAAGCCAGATGCCCCCA 15076  
 QY 1851 sGlyTyrThrCysGluCysProProAsnTyrLeuGlyProTyrCysGlyThr----- 1868  
 Db 15077 TGGCTATACCTGCGAGTGTCCCAAAATACCTTGGGCCATATGTGAGACAGTAAAC 15136  
 QY 1868 ----- 1868  
 Db 15137 AGACAGGGCATGTGGAGAGAGTCCAGTGGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 15196  
 QY 1868 ----- 1868  
 Db 15197 GAGCCCGAAGCCTGCTGATTCACAGACAGAGGTCAAGAGGCAACATAGGGCTACCT 15256  
 QY 1869 -----ArgIleA 1871  
 Db 15257 AGTTAGGTGAGTGGCAGGGGAGCTCATGCTTACCTGGGCTCTCTGCAAGATTG 15316  
 QY 1871 sPglNProCysPArgGlyTyrTrpGlyNhiAspProThrCysGlyProCysAsnCysAspV 1891  
 Db 15317 ACCAGCCTTGTCCCGGTGGCTGTGGTGGGACATCCACATGTGGCCATGCAACTGTATG 15376  
 QY 1891 a1SerIleGlyPheAspProAspProAspNhiSerGlyGlyCysHisCysIleS----- 1909  
 Db 15377 TCAGCAAAAGCTTGAACCAAGCTGCAACAAAGAACAGCGGAGTGCACCTGCAAGGTGA 15436  
 QY 1909 ----- 1909  
 Db 15437 CAGCCCCAAGCAGCCTTCCACTGTGSCCATTTGGGCTCTGTGCCATCTCTGGGCCCT 15496  
 QY 1909 ----- 1909  
 Db 15497 AGAGGGCAGGCTTAGAGGATGGCTTCTCTGTAGAGGTGAGGGCTTGTGCTTGTGCT 15556  
 QY 1909 ----- 1909  
 Db 15557 CACAGCAGGGCAGTGTGCTTACCTGTGTGTCTGAGGCGCAGGGGAGCCCTGGGC 15616  
 QY 1909 ----- 1909  
 Db 15617 CAATGTGAACCATTAAGCTCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 15676



Db 17837 TGAGAACGGGGCTTCTGGAATTCAGCCTGTGTCTTCTGGGCTCCAGCTGAGAGGCC 17896  
 QY 2036 -----  
 Db 17897 GTCTCTACCTCTGAGGCTCTTTTGCTCCAGAGTTCTGCCCTCACTCTGCTCTCCG 17956  
 QY 2037 -----  
 Db 17957 CTCGGCTGTGTAGGCTGAGCGGCTACAGCGGATAGTCAAGGCTTAGACTCAGGGCGC 18016  
 QY 2052 SerGlnGlnLeuAlaLeuLeuLeuArgAnaIaThrGlnHisPheValaGlyPheGly 2071  
 Db 18017 TCCAGACGCTACCCCTGCTCTCTGCGCAAGCCAGCAGCAGCAGCAGCTGACTTCGAGC 18076  
 QY 2072 SerAspValIleValaIaTyrgIleuAlaThrArgLeuLeuAlaHisGluSerThrGln 2091  
 Db 18077 AGGAGCTCAAGGTGGCTTACCAAGCTGGCCAGCGGCTGTCTGCCCCACAGAGCACCCAG 18136  
 QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlu----- 2106  
 Db 18137 CGGGGCTTTGGGCTGTCTGTCCACACAGGACGTGCACTTCACTGAGGGTGGGGCTTGGAGGA 18196  
 QY 2106 -----  
 Db 18197 TGCAGGCTGCGCTGGTTAGATAGGGGTCA TGGTAGTGAAGCTGCTCATGAGCAACTGGG 18256  
 QY 2106 -----  
 Db 18257 GGGCAGAGGGGGCCCTCCATCCACCTACCAAGAGGCTCCGTGTGGCAGAGGCTCCA 18316  
 QY 2106 -----  
 Db 18317 TGAACCCCTGTGACCCCTGAGCCCTCTTCCAACCTTCTCATAGTAACATATGNG 18376  
 QY 2106 -----  
 Db 18377 TGTGGGGGATGATGTGACCCAGAGCAGAGGCTGCTGTGTACCAATGGGCTGGAGGA 18436  
 QY 2106 -----  
 Db 18437 AGCAGTATCTCAGGACATGTGTGTGGGTCCAGGCAAGGGCTGGAAAGCTTATGTAGA 18496  
 QY 2106 -----  
 Db 18497 GAATGAGAGGAGGGCTTAGGGGCAAGACATCAGTGAACCTGATGGGCAGTGCGGTCCA 18556  
 QY 2106 -----  
 Db 18557 GGAGCATTTCAAGGGAGGCAAGTGAACAAGAGGAGGGGCTGTGTGGGAGCAAGGA 18616  
 QY 2106 -----  
 Db 18617 CTGGCCCGGGGCAAGAGCGGGGCTCATGTGTGGGATGTCAAGGTGTGGTTGTAGACA 18676  
 QY 2107 -----  
 Db 18677 CCCACCGTGAACCTTGCACCCAGAAATCTGCTGGGGGTGGGCAAGCGCTCTGAGAC 18736  
 QY 2118 TAlaAsnIleValArgHisTTPGLeuLeuIleGlnGlnThrGluGlyValaIleuLeuAspTh 2138  
 Db 18737 AGCCCAACAGCGGCACTGGAGCTATCTCAGGACAGAGGGGTGGCACCGCTGTGCTCT 18796  
 QY 2138 uGlnHisTyrgIleuAlaTyrgIleuAlaLeuAlaGlnAsnMetArgHisTThTyLeuSe 2158  
 Db 18797 CCAGCACTATAGGGCTACGCGCAGTGCCTGGCCCAAGAACATGGGGCAACCTTACTAAG 18856  
 QY 2158 rProPheThrIleValIleThrProAsnIle----- 2167  
 Db 18857 CCCCCTTCAACATGTCACGCCCAACATTGGTAAGGCTGTGCTTGGGTTGGGAGGGGTT 18916  
 QY 2167 -----  
 Db 18917 TGTGAGAGGAGTCCCCGACAGAGCGGCTGTGTGGGGTCTGTGCTGCTCAAGGTCCG 18976

QY 2168 -----  
 Db 18977 ATCTGTGACATCCCCCTTCTTAGTCACTCTCGTGTAGCGCTTGGACAAAGGAACCTTTG 19036  
 QY 2180 IagIValIleValLeuProArgTyrgIleuAlaLeuArgIleuGlnProProAspLeuGluT 2200  
 Db 19037 CTGGGGCCAGCTGCGCCGCTACAGAGGCCCTGTGCGGGAGAGCAGCCCGGACCTTGAGA 19096  
 QY 2200 hThrValIleLeuProGluSerValPheArg----- 2210  
 Db 19097 CAACGATTTCTGTGCTGATGTGTCTTCAAG-AGGTCAGTGTGTGACATGATTAAGTTG 19155  
 QY 2210 -----  
 Db 19156 GGAAGCTGAGACCCAGTGTCTGTGCAAGACTCCACAGAGAGCAGGGGCCAGTAAAGTGAAC 19215  
 QY 2211 -----  
 Db 19216 AGTGTCCCTCCAGAGAGAGCCCGGTGTCAAGGCCCGCAGAGGCCCGGAGAGGCCCAGG 19275  
 QY 2226 IuProGluGlnLeuAlaArgArgIleuArgHisPProGluLeuSerGlnIleuValaVal 2246  
 Db 19276 AGCCAGAGAGCTGGCAGCGGCAAGCGACGCGACCCGAGCTGAGCCAGGGGTAGGCTG 19335  
 QY 2246 aAlaSerValIleIleTyArgThrLeuAlaGlyLeuLeuProHisAsnTyraAspProA 2266  
 Db 19336 TGGCAGGCTCATCATCTACCGCAACCTGGCCGGGCTACTGCTCATTAATGACCTTG 19395  
 QY 2266 sPlyArgSerLeu----- 2270  
 Db 19396 ACAAGCGCAGCTTGAAGGTCAAGCAGCTAGGGGACAGAGTGTGTAGGGTATAGGTGCGGAC 19455  
 QY 2270 -----  
 Db 19456 GGTAGTGTGAGGCAATGAGAGGGGTGCGGGGCTCTCCCAATGATGATGCTGCGGTG 19515  
 QY 2271 -----  
 Db 19516 TGACGTGCACTGACCTGACCTGCCCCACATGAGAGTCCCAAAACCCGATATACACACACC 19575  
 QY 2282 ValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspIlePro 2301  
 Db 19576 GTGTGTGACATCAACGCTTCAATGATGAGAGGAGCTTCTGCCCCGGGCTTGGACAAACCC 19635  
 QY 2302 ValThrValGlnPheArgLeuLeuGluThrGlnIleuArgHisPheValPhe 2321  
 Db 19636 GTCAAGGTGCAGTTCGCGCTGTGTGAGACAGAGGAGGAGACCAAGCCCATCTGTGTCTTC 19695  
 QY 2322 TrpAsnHisSerIleLeu----- 2327  
 Db 19696 TGAACCATTCATCTCTGTAGACCTGCACTGCGCCCTGAGGCTTCGGGCTGAAA 19755  
 QY 2327 -----  
 Db 19756 GTCCAGGCCCCGTGACATCAACCTCTGTCTCCCTGACCTTGCCTTCTCACAGAG 19815  
 QY 2328 ValSerGlyThrGlyValTyTPSerAlaArgIleCysGlyValValPheArgAsnGluSer 2347  
 Db 19816 GTCAGTGCAACAGGTGTGTGTGTGCGCCAGAGGCTGTAAAGTGTCTTCCGAAAGAGAGC 19875  
 QY 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArg 2367  
 Db 19876 CAGGTCAAGCTGCGCAAGTCAACACATACAGAGCTTGTGCTGCTCAAGGAGCTTCTCGG 19935  
 QY 2368 ArgGlu----- 2369  
 Db 19936 CGGAGGTGCGGGCCACAGAGGGGAGCTGCAAGGCGGTGGTGGGACCCAGAGGACCGGG 19995  
 QY 2370 -----  
 Db 19996 CTGGGTGCTAGGTCCTGCTTCTTAATTCCTGTGGCCCTTGCACACTACTGTGAGAA 20055





<http://www.sanger.ac.uk/HGP/Chr1>  
RP11-29704 is from the library RPC1-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6  
This sequence is the entire insert of clone RP11-29704. The true  
left end of clone RP11-173K24 is at 98796 in this sequence. The  
true right end of clone RP11-352P4 is at 101180 in this sequence.  
location/Qualifiers

## FEATURES

```
source
1..169241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-29704"
/clone_1fb="RPC1-11.2"
1676..11775
repeat_region
/note="MIR repeat: matches 45..152 of consensus"
2574..3410
misc_feature
/note="CpG island"
evidence=not_experimental
2809..2936
repeat_region
/note="64 copies 2 mer gg 60% conserved"
2814..2943
repeat_region
/note="13 copies 10 mer ggcggggcg 63% conserved"
2982..2987
misc_feature
/note="Sequence from overlapping clones BA352P4 (
AL356389) and d641D22 (AL157901). Assembly confirmed by
restriction digest."
3007..3116
repeat_region
/note="11 copies 10 mer gcccgggcg 63% conserved"
3364..3411
repeat_region
/note="16 copies 3 mer tgc 83% conserved"
9894..10207
repeat_region
/note="AluSg repeat: matches 1..296 of consensus"
10332..10617
repeat_region
/note="AluJ repeat: matches 10..299 of consensus"
11683..12328
misc_feature
/note="CpG island"
evidence=not_experimental
12551..12727
repeat_region
/note="MIR repeat: matches 10..197 of consensus"
13355..13442
repeat_region
/note="44 copies 2 mer gg 65% conserved"
15405..15448
repeat_region
/note="22 copies 2 mer gt 75% conserved"
17917..17995
repeat_region
/note="MIR repeat: matches 45..129 of consensus"
19254..19293
repeat_region
/note="4 copies 10 mer aggcctggcg 95% conserved"
19651..19956
repeat_region
/note="AluSg repeat: matches 5..309 of consensus"
19972..20270
repeat_region
/note="AluSc repeat: matches 1..296 of consensus"
26166..26195
repeat_region
/note="10 copies 3 mer gag 90% conserved"
29612..29726
repeat_region
/note="FLAM C repeat: matches 1..118 of consensus"
29778..30077
repeat_region
/note="AluSx repeat: matches 1..298 of consensus"
30135..30428
repeat_region
/note="AluJb repeat: matches 1..298 of consensus"
30454..30616
repeat_region
/note="MIR repeat: matches 98..256 of consensus"
32159..32470
repeat_region
/note="AluSd repeat: matches 1..312 of consensus"
32481..32780
repeat_region
/note="AluSx repeat: matches 1..298 of consensus"
34670..34803
repeat_region
/note="MIR repeat: matches 123..250 of consensus"
36339..36835
misc_feature
/note="CpG island"
evidence=not_experimental
```

```
repeat_region
37313..37937
/note="MER4B repeat: matches 1..635 of consensus"
37938..38051
repeat_region
/note="MER4B repeat: matches 103..213 of consensus"
38052..38351
repeat_region
/note="AluY repeat: matches 1..300 of consensus"
38352..38753
repeat_region
/note="MER4B repeat: matches 213..635 of consensus"
38987..39182
repeat_region
/note="MER4B repeat: matches 1..171 of consensus"
39183..39482
repeat_region
/note="AluSx repeat: matches 1..300 of consensus"
39483..39504
repeat_region
/note="MER4B repeat: matches 170..191 of consensus"
39537..39824
repeat_region
/note="AluSx repeat: matches 1..304 of consensus"
39850..40383
repeat_region
/note="MER4A repeat: matches 1..554 of consensus"
40394..40483
repeat_region
/note="MER4D repeat: matches 359..448 of consensus"
40532..40838
repeat_region
/note="AluY repeat: matches 1..311 of consensus"
40956..41169
repeat_region
/note="L1P1 repeat: matches 5927..6151 of consensus"
41170..41472
repeat_region
/note="MER4D repeat: matches 433..745 of consensus"
41473..41771
repeat_region
/note="AluY repeat: matches 2..300 of consensus"
41772..42011
repeat_region
/note="MER4D repeat: matches 745..981 of consensus"
42399..42702
repeat_region
/note="AluSx repeat: matches 1..304 of consensus"
42948..43256
repeat_region
/note="AluY repeat: matches 1..310 of consensus"
44024..44141
repeat_region
/note="MIR repeat: matches 21..147 of consensus"
44187..44622
repeat_region
/note="MIR repeat: matches 77..547 of consensus"
44790..45097
repeat_region
/note="AluSg repeat: matches 1..306 of consensus"
45098..45232
repeat_region
/note="AluSg/x repeat: matches 1..136 of consensus"
46712..47013
repeat_region
/note="AluY repeat: matches 1..305 of consensus"
47032..47323
repeat_region
/note="AluSx repeat: matches 1..292 of consensus"
47669..47826
repeat_region
/note="L2 repeat: matches 2545..2709 of consensus"
48017..48204
repeat_region
/note="MIR repeat: matches 22..209 of consensus"
49692..49988
repeat_region
/note="AluSx repeat: matches 1..297 of consensus"
54709..55019
repeat_region
/note="AluSg repeat: matches 1..311 of consensus"
55261..55380
repeat_region
/note="60 copies 2 mer gt 79% conserved"
55267..55386
repeat_region
/note="12 copies 10 mer gtgctgtgc 79% conserved"
55386..55419
repeat_region
/note="L2 repeat: matches 330..834 of consensus"
55756..56029
repeat_region
/note="L2 repeat: matches 46..328 of consensus"
56258..56327
repeat_region
/note="MIR repeat: matches 323..388 of consensus"
56425..56915
repeat_region
/note="L2 repeat: matches 330..834 of consensus"
56926..57541
repeat_region
/note="L2 repeat: matches 2097..2748 of consensus"
58447..58523
repeat_region
/note="MIR repeat: matches 49..146 of consensus"
59152..59300
repeat_region
/note="MIR repeat: matches 29..185 of consensus"
60773..61014
```



```
/note="MIR repeat: matches 4. .262 of consensus"
repeat_region 61115. .61227
/note="MIR repeat: matches 36. .150 of consensus"
repeat_region 61443. .61519
/note="MIR repeat: matches 6234. .6314 of consensus"
repeat_region 61956. .62063
/note="MIR repeat: matches 35. .141 of consensus"
repeat_region 63402. .63437
/note="18 copies 2 mer ta 80% conserved"
repeat_region 68939. .69310
/note="Alu repeat: matches 2. .306 of consensus"
repeat_region 69476. .69781
/note="Alu repeat: matches 1. .292 of consensus"
repeat_region 70577. .70718
/note="MIR repeat: matches 2. .155 of consensus"
repeat_region 71406. .71699

Alignment Scores:
Pred. No.: 0 Length: 169241
Score: 12339.00 Matches: 2907
Percent Similarity: 36.92% Conservative: 0
Best Local Similarity: 36.92% Mismatches: 8
Query Match: 9 Indels: 4967
DB: Gaps: 32

US-09-916-849a-3 (1-2923) x AL390252 (1-169241)
QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20
DB 3330 ATGGGAGAGCCGGGACCGGGGCTCCCTCCCAAGCCGCGCCGCTCTGCTG 3389
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSer 40
DB 3390 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyTyrPheCysProSer 60
DB 3450 GGGTTCAGAGGAGCGAGGCTCTTGGGGGCTGCGCCCGCATGGGCTGCTGCTCA 3509
QY 61 SerAlaSerAsnLeuTyrPheTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80
DB 3510 TCAGGCTGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3569
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100
DB 3570 GGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3629
QY 101 IleProLeuProProAlaProGlnGlyCysProTyrPheCysArgLeuLeuGlyTyrGly 120
DB 3630 ATTCCTTACCAACGAGCTCTGAGAGCTGCTGCTGAGCTGCTGCTGCTGCTGCT 3689
QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGlnHisProCysLeuValAla 140
DB 3690 GGGCACTTTCCTCCAGGAGGAGCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 3749
QY 141 ProArgLeuValGlyGlnSerCysLeuValAlaGlnAlaProGlyLeuValGlyGln 160
DB 3750 CCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3809
QY 161 ArgSerProGlnGlnSerLeuGlyGlyArgArgLeuValAlaMetThrAlaProGln 180
DB 3810 AGGTCAACCAAGAGTCTCTGGGGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 3869
QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200
DB 3870 TTCAGAGCCCTCCAGCTACCAAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 3929
QY 201 AlaSerLeuValAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlyTyrThrMet 220
DB 3930 GCATCTCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3989
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
```

```
DB 3990 GATGCCCTCTTTGATGACCGCTCCACAGTCTTCTCCCTGAGCCCACTCATGTGCA 4049
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyThrIleSerThrHisValPheArgValThr 260
DB 4050 GTTACCAACGAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4109
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280
DB 4110 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4169
QY 281 AspThrAspHisAspAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 4170 GACACCAATGACCAATGACCTGTGTGTGACGAGGAGGAGGAGGAGGAGGAGGAG 4229
QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320
DB 4230 AACCTGAGGAGTGGCTGAGGAGTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 4289
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340
DB 4290 AATGCCAATATTCGTACCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4349
QY 341 GlnIleAspProAspSerGlyValIleArgThrArgGlyProValAspArgGlnGlnVal 360
DB 4350 GAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4409
QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380
DB 4410 GAATCTTACCAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4469
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAspAspAlaProGlnPheSer 400
DB 4470 ACCAAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4529
QY 401 GlnIleAspArgTyrValGlnValArgGlnAspValThrProGlyAlaProValLeuArg 420
DB 4530 GAGAGGCTTATATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4589
QY 421 ValThrAlaSerAspAspAspAspGlySerAsnAlaValAlaIleTyrSerIleMetSer 440
DB 4590 GTTCAAGCTCTGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4469
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
DB 4650 GGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4709
QY 461 ProLeuAspTyrGlnThrThrIleGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480
DB 4710 CTTCTTGAATATGAGACGACCAAGAGTACCTTACGGGTGCTGAGCAGAGTGTGG 4769
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
DB 4770 GGTCCCACTCTCTATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4829
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520
DB 4830 AATGCCCAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4889
QY 521 GlyTyrLeuValLeuHisValAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
DB 4890 GGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4949
QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly 560
DB 4950 GAATACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5009
QY 561 TrpIleSerValAlaAlaGlnLeuAspAspGlnGlnValAspPheTyrSerPheGlyVal 580
DB 5010 TGGATCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5069
QY 581 GlnAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValThrVal 600
DB 5070 GAGCTGAGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5129
```

QY	601	LeuSpApValAaSpApAaSPProThTh-PheThcGlnPProGluYrTrThrValAArgLeuAaSn	620
Db	5130	CTGAATGTCACAGCAACATCAATCCATTACCCCAACAGATGACAGTGGGCTCAAT	5189
QY	621	GluApAValAaValGlyThrSerValAThrValSerAlaValAaParGAspAlaHis	640
Db	5190	GAGATGACAGCTGTGGGACCAAGGTGGGTGACGGGTGTCAGCTGTGGACCGGTGATGCTCAT	5249
QY	641	SerValAlIeThrTyrrGlnIleThrSerGlyAsnThrArgAaAaGPhnSerIleThrSer	660
Db	5250	AGTGTCAATCACTTACCAAGATCAACAGTGGCAATCTCGAAACCGCTTCTCATCAACCAAGC	5309
QY	661	GlnSerGlyGlyGlyLeuValSerIleuAlaLeuPProLeuAaPyrTyLLeuGluAaGln	680
Db	5310	CAAGGTGGTGGGGCTGTATCTCTTCCCTCGCACTGAGACTACAACTTGAAGGGCAG	5369
QY	681	TyrValLeuAlaValAThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValAl	700
Db	5370	TATGTGTGGCTGTACCGCTCCGATGGCACTGGGACGAGCAACGGACAGATTGTGGT	5429
QY	701	AsnValAThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrrThValAaSn	720
Db	5430	AATGTCAACGACGGCCAAACCACTGCTCTGTTCAGAGCTCCCATATACAGTGAAT	5489
QY	721	ValAaGlnAaParGProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAaP	740
Db	5490	GTTTATATAGAGACCGGGCCGGACGGACCAACGGTGTGCTGATCAAGCCACCAAGAGAAC	5549
QY	741	ThrGlyGluAaAaAlaArgIleThrTyrrPheMetGluAaSerIleProGlnPheArgIle	760
Db	5550	ACAGGTAGAGATGCCCCGATCACCCTACTTCAATGAGGAGCAACATCCCCAGTCCGATC	5609
QY	761	AspAlaAspThrGlyAlaValAThrThcGlnAlaGluAaAaPyrGluAaGlnValSer	780
Db	5510	GATCAACAGCAAGGGGGCTGTCAACCAACGAGCTGAGCTGAGCTATGAAGACCAAGTGTCT	5669
QY	781	TyrrThrLeuAlaIleThrAlaArgAspAaGlnGlyIleProGlnTyLSerAspThrThrTyrr	800
Db	5670	TACACCTCGGCACTTACTGCTCTCGGACAAATGGCATTCCCCAGAACTCCGACCACTCAC	5729
QY	801	LeuGlnIleLeuValAaSnAspValAaSnAaAaAaAlaProGlnPheLeuArgAspSerTyrr	820
Db	5730	CTGAGAGATCCCTGGTGAACGATGAAATGACATGCCCCCTCAGTTCCTCGAGACTCTCAC	5789
QY	821	GlnGlySerValTyrrGluAaPValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	5790	CAGGCAAGTGTCTATGAGAAATGTGCACCCCTTCACTAACCGTCCGCAATCTCAAGCACT	5849
QY	841	AspArgAspSerGlyLeuAaGlnGlyAArgValPheTyrrThrPheGlnGlyGlyAAspAaGly	860
Db	5850	GATGTGATTTGTGACCTTAATGGCAAGGCTCTTCTACACCTTCCAAAGAGGGGACATATGA	5909
QY	861	AspGlyAaPheIleValGluSerThrSerGlyIleValAArgThrLeuArgArgLeuAaP	880
Db	5910	GACGGTGACTTATATGTTGAGTTCACCGTCAAGCAATCGTGCAACGCTTCGAGAGGCTGAT	5969
QY	881	ArgGluAaSnValAlaGlnTyrrValLeuArgAlaTyrrAlaValAaAaPlyGlyMetProPro	900
Db	5970	CGAAGAAACGGGCCCAAGTATGTCTTGGGGCATATGACATGAGCAACAGGAGATGCCCCCA	6029
QY	901	AlaArgThrProMetGluValAThrValAThrValLeuAaPValAaSnAaAaSnProProVal	920
Db	6030	GCCCGACACCTATATGAGATGACATCACTGTGTTGATATGATATGACAAATCCCTCTCTC	6089
QY	921	PheGlnGluAaPgluPheAspValPheValGluGluAaSnSerProIleGlyLeuAlaVal	940
Db	6090	TTTATGACGAGATGATGTTGATATGTGTTTGTGAAAGAGAAACACCCCATGGGCTAGCCGTG	6149
QY	941	AlaArgValAThrAlaThrAspProAspGlnGlyThrAaAaAaGlnIleMetTyrrGlnIle	960
Db	6150	GCCCCGGTCAACGACCTGACCCCGGATGAAGCAACAAATGCCAGATTATGTATCAACGATT	6209

QY	96	ValGluIuYAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla	1980
Db	6210	GTGGAGGGCAACATCCCTGAGGCTCTTCACTGGACATCTTCCGGGAGCTACAGCC	6269
QY	981	LeuValAspLeuAspIlyrGlyuAspArgProGluValLeuValIleGluIleThrSer	1000
Db	6270	CTGGTAACTTAAGCTACAGAGACCAGGCTGAGTACGCTCGGTCACTCAAGGCCACTCA	6329
QY	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro	1020
Db	6330	GCTCTCTGGAGGCCGGGCTACAGTCCAGTCCGGCTCTTGACCCGCAATGACACACCA	6389
QY	1021	ProValLeuGlyAsnPheGluIleLeuPheAsnAsnIlyrValThrAsnArgSerSer	1040
Db	6390	CCAAGTGGGCAACTTGAAGATCTTTCAACAACATATGATCAATTCGCTCAAGCAGC	6449
QY	1041	PheProGlyGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	6450	TTCCCTGGGGGTGCATATGGCCGAGTACCTGCCCATACCTGATATCTCAGATAGCTG	6509
QY	1061	ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080
Db	6510	ACTTACAGCTTTGAGCGGGAAATGAATCAAGCTGTCTGTCTCAATGCTCCAAGGT	6555
QY	1081	GluLeuIlySerLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	6570	GAGCTGAAGCTAAAGCCCGCACTGACAAACAACCGGCTCTGGAGGCAATCATAGCGTG	6629
QY	1101	LeuValSer-----	1103
Db	6630	CTGGGTGTC-AGTAAAGAGAGGCCCAAGGTGGCGCTGGGGGTAGCTCGCGGGATG	6688
QY	1103	-----	1103
Db	6689	GTCTGGGAGAGCACTGAGAGGAGGTGCATCCAGAAAGACTACAGATCCACTCCCT	6748
QY	1103	-----	1103
Db	6749	GCCCAAGTCTGAGCAGAGAGAGGGGCAAGACCAAGCTTGGAAAGAACCCCAAGAA	6808
QY	1103	-----	1103
Db	6809	CCCTGGGGGCTGTGTGACAGGCCCACTTGTCTGGGCTGGGCTGGGCTTGGGTTGGC	6868
QY	1103	-----	1103
Db	6869	TGCCCAACTGTCTACAGCCTGTGACGTGTGGGAAAGTGTGTGAGAGCTGCCCGGC	6928
QY	1103	-----	1103
Db	6929	TATGGCTGCCAAGATTGTGAAAAAGCGCGTGAAGCTGGGGGGCCAGAGAGCTGG	6988
QY	1103	-----	1103
Db	6989	GGAAAGGGCAGAGAGACTGTGTGGCAGAGAGAGGCGAGGCCAGGCCCTTCACCCC	7048
QY	1103	-----	1103
Db	7049	CATCAAGATATTCTGACGTCTCAACAGACCTTTGGAGAGCTGGGACCAAGGAGAC	7108
QY	1103	-----	1103
Db	7109	AGGAGTGCAGCAACCAAGGGGTACAGAACCTGTATGTCTCTTGTGGCCCTTTTGTCTCC	7168
QY	1103	-----	1103
Db	7169	TGGCTGAGCTGCAGAAATCGCTGACAGGGACTTGTGTACGGCAATCTTTCCCTCAGG	7228
QY	1103	-----	1103
Db	7229	CAGAAAGAGCTGGCTGGCTCCCTCCCTTCAGAGCTCCCAAGAAAGTGAAGCCAGAGCTC	7288
QY	1103	-----	1103

Db 7289 ATTGTCTGCAAAACAGACCCCACTTTCAGAGTCCGTGCTGGGTCTGCTTGAGAG 7348  
 QY 1103 ----- 1103  
 Db 7349 GGGCTGATGAAGCTGGGGGCGAGGCTGAGGCTGTGGGCCCCCTTGCCACTCATG 7408  
 QY 1103 ----- 1103  
 Db 7409 GGCCTTCTTAGGGTGCAGGCTTCTTCTGCTAGCTGTGACCAAGCCAAAGCTCCAG 7468  
 QY 1103 ----- 1103  
 Db 7469 GAATGAGTGTAGGACCAAGACAGGCCAGTGAAGTGTGAGCTTTGAGGGAGAGACC 7528  
 QY 1103 ----- 1103  
 Db 7529 CAAGGACCTTATCTGAGAGAGAACAGAGTCCAGGGGCTGGAGCTTGCCGGAACCC 7588  
 QY 1103 ----- 1103  
 Db 7589 AGGCTGGGCAATTCCTGCTGTGCGCAAGCACTCAGCTTACAGACTCCCTGAGACAG 7648  
 QY 1103 ----- 1103  
 Db 7649 GGCTGGGGAGAGAGTTGAGAGGGGTGTGACAGTATTCGTCTTTTATACAGCTG 7708  
 QY 1103 ----- 1103  
 Db 7709 TGGGAGCCAGAGTACAGAGGGAGAAACACCAAGTTGCTGAGTCCCTTCCC 7768  
 QY 1103 ----- 1103  
 Db 7769 AAGCTGGAGCACTGAGACCCCTATGAGGGAGTGTGTAGTGCACCCAGAGGGTGG 7828  
 QY 1103 ----- 1103  
 Db 7829 ATCAGTCAATCCCTCTAGGAGCTTCAATCCATCAGGAAAGGAGAAAGAGA 7888  
 QY 1103 ----- 1103  
 Db 7889 GGCTGGAGGCCCCCTAATCCACCCGAGGAGTGTGGCTTGAAGGGAGAGAGGGTATG 7948  
 QY 1103 ----- 1103  
 Db 7949 CCAGAGTCAGGGCTGAGCTCACTCAGGGTCTTGTCTGAGCAGAGCTCCTGCCAC 8008  
 QY 1103 ----- 1103  
 Db 8009 CCTTCAAAAGTACCAACAGAACCAAGAGCAGAGCTTTCAAGGACTGTGGCTGAGTT 8068  
 QY 1103 ----- 1103  
 Db 8069 CTGAGGCTCTGAGACTTCAAGTTTCCCTCAAGAGGGCGGAGAGAGGCTGAGCTCC 8128  
 QY 1103 ----- 1103  
 Db 8129 TGTGTCTCACTGTGCTGAGCACTTGAGACAGTGTGAGAGAGGTGGAGAGAGGCGC 8188  
 QY 1103 ----- 1103  
 Db 8189 TGTCCCTGTGTTCTCAGACCCAGGCTGCTTTGTGTAGGCTGGCCAGGCACTGGCAGA 8248  
 QY 1103 ----- 1103  
 Db 8249 GAAGGCCCCCATTAAGCTTGGGTCTCCACAGGGGAGAGCCCTAAGACTTGGTTCC 8308  
 QY 1103 ----- 1103  
 Db 8309 TAGTCTAAGAGAGCTCGGGGCAACATCCAGAGCGGCTCTCTGAGCCGCGAGCTCTGG 8368  
 QY 1103 ----- 1103

Db 8369 CCCACCTCTCCCTTGGGCTTCAGCCTTGAGGCACTCAACAGCAACCCCAAGT 8428  
 QY 1103 ----- 1103  
 Db 8429 CAGGAGGCGGGGTGAGGGGAGCTGACTCCACTGCCCTTTGTTTCTCTCTTGT 8488  
 QY 1103 ----- 1103  
 Db 8489 TCTTGTCTTCTTCTTCTTCCCGCCCTGCCATGCGGAGCCTGACACTGCCCTTC 8548  
 QY 1103 ----- 1103  
 Db 8549 TCCGTGCCCTCTCCCTGCTCTGAGGCTCTCGGCTCTCCACTTGCCATTGCTGCT 8608  
 QY 1103 ----- 1103  
 Db 8609 CCTCTCTGAGGCTTGTCTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8668  
 QY 1103 ----- 1103  
 Db 8669 CCTATGCTGAGCTGCCCGAGCCTGGGCGATGGCTCCCTGAGGCTGAATTCCTTC 8728  
 QY 1103 ----- 1103  
 Db 8729 GCTCTTTCTTCTCTGCTCAGACACGACAGAGCTCTGTTCATGTCTGGGAGCTCC 8788  
 QY 1103 ----- 1103  
 Db 8789 AGAACAGAGAGCCCTGTGGCTTCAGTCTCCGCTGACATCTCCAGGGAGAGAG 8848  
 QY 1103 ----- 1103  
 Db 8849 GCTGGAGCTGACCGGAGCTGCACTTTCAAGTGTGCTGGCTAAGCTTGGACAC 8908  
 QY 1103 ----- 1103  
 Db 8909 AGTGGCATTTCTTGGGCTCAGAGGCCATCCAGCTGTTCTTTTGTGTCCAGAAACA 8968  
 QY 1103 ----- 1103  
 Db 8969 GCTAGGCAAGTACTCCCAACCCCTGTTCCTTCATCTGCTTGTGACCTGAGGGTAC 9028  
 QY 1103 ----- 1103  
 Db 9029 TAGTCAAGCAGGGGAGGTGATATAGAAAGAGAGCAGTGGTTGACTGTTGGGA 9088  
 QY 1103 ----- 1103  
 Db 9089 TGGGAGACACAGACAGAGGGGCAAGTGGCCAGATGTGTGACCAATGGAAACACA 9148  
 QY 1103 ----- 1103  
 Db 9149 GAGCATGTGTGGAGAGAGGGGCCAGGGGCTTAAGACATGGCGGTGGGGACACAG 9208  
 QY 1103 ----- 1103  
 Db 9209 GAGAGTGTGTTCCAGAGCTGAGGGGTGGCAAGTTCTCCGCTCATGSCCTTG 9268  
 QY 1103 ----- 1103  
 Db 9269 TTGCCCCCTCTCTCTTATATCAATTATGGGCACTGTGTGTGAGTCTTGGGCTGTT 9328  
 QY 1103 ----- 1103  
 Db 9329 CTGCTGAGGGAATATCCGTCAAGAAATCATTAAGGAGATATACAGGTATACAGAG 9388  
 QY 1103 ----- 1103  
 Db 9389 CTGTGTCAAGTGAATCTGTGTGCGGGAGTCTGAACGAGAGAAAGTAGACC 9448  
 QY 1103 ----- 1103  
 Db 9449 TGATGACATAAATCTGGGCAAGCTTCTTGAGAGCAGTGTGAGGTAGGGCAGTTAAGACT 9508

QY 1103 ----- 1103  
Db 9509 CAGAGACTGTGGCTGTGTCTGTGGTGGAAGAACGCTGTGCAGCAGCTTTGAGAC 9568  
QY 1103 ----- 1103  
Db 9569 GAGAGCAGAGAAACAGAGTGGGTATGACAGCAGGCTTGTGTGAGAGCTAGTAG 9628  
QY 1103 ----- 1103  
Db 9629 AATTAAGCAATGATGAGGCTGAGGGTAGGTGGCAGTCCGCTGAGAGCATGTGAGAT 9688  
QY 1103 ----- 1103  
Db 9689 GGGGCTGTCTTCAAGCTTCCGGCTTCAAGTGAAGCCAGCTGGTAAGTCAAGCACC 9748  
QY 1103 ----- 1103  
Db 9749 GCAAGTAGTACACACAGATGCAATTTGTTCCATGTCCCTGTGCTACTCCACCCC 9808  
QY 1103 ----- 1103  
Db 9809 ACCCAGAAACATGGGCTCCTGCCGTACTGTAGGGTAGCAATTGCTTTTATATC 9868  
QY 1103 ----- 1103  
Db 9869 TAGTTGGGCTCCCCCTGTATGTGTGTTTGTGTTTGGAGACAGATCTCACTCTG 9928  
QY 1103 ----- 1103  
Db 9929 TCACCAAGCTGAGTACATATGGGCGATCTGGCTCACTGCAACCTCTGCTGGGT 9988  
QY 1103 ----- 1103  
Db 9989 TCAGCAATTTCTCTGCTCAGCCTCTGTAGTATCTGGAGCTCTGAGTAGAGATT 10048  
QY 1103 ----- 1103  
Db 10049 ACAGGACCCACCAAGCCCACTAATTTTGTATTTTATAGTAGAGAGGGGTTTCA 10108  
QY 1103 ----- 1103  
Db 10109 CATGTGTGAGTCTGATCTGAACTCTGACTGTGATCCACCGGCTCAAGCTCCCA 10168  
QY 1103 ----- 1103  
Db 10169 AAGTTCTGGACTACAGGCGTAGCCACTGCGCCCGGCTGTGTATGTTTAAAGTCC 10228  
QY 1103 ----- 1103  
Db 10229 GTTTATTTCTTGTGAGACATGAATAGATTCACATCTGAGTCTTTCTGAACACC 10288  
QY 1103 ----- 1103  
Db 10289 TACTATGTGAAGCATTTGTATTAATAAATAATGAATATGCAAGTAGCTATGCTGT 10348  
QY 1103 ----- 1103  
Db 10349 AATCCAGTGTCTTAAGAGGCTGAGATGAGATCACTTGAAGCCAGAAATTTGAGACCA 10408  
QY 1103 ----- 1103  
Db 10409 GCCTGACAAACAGCAAGACCCGATCTTACACACAAACAAAATTAAGCAGGCAATGT 10468  
QY 1103 ----- 1103  
Db 10469 TGTGAGCACTGTGTCTCAGCTATTTGGAGTTGAGTGGAGGATTCGCTTGAAGCCA 10528  
QY 1103 ----- 1103  
Db 10529 GGAAGTTGAGGCTACAGTAGCTATGATCTGCCACTGTCTCCAGTGTAGGCGACAGAG 10588

QY 1103 ----- 1103  
Db 10589 AGTCTGTCTTAAAAAAGACTATCCCTTCTTGGCTGTAAAGCAATT 10648  
QY 1103 ----- 1103  
Db 10649 CCTCTTCCAGGCCCGAGCTTTCATTTTATAGCTAATAAATTCATTCCTATGTGA 10708  
QY 1103 ----- 1103  
Db 10709 TGTGATTCAGGTGTATGTATTTCTGGCTTAGGGTGTGTGGTAGATGAGACACC 10768  
QY 1103 ----- 1103  
Db 10769 AGCAAGTGTATTAGTGTAGCTTATAGTAAAGTAGTGTTCCTGGGGCATAGCA 10828  
QY 1103 ----- 1103  
Db 10829 GGCATTTGGTGTATATGTTAGGCCAAGTGGTAGATGACTAGAGCCCTAGAGCCA 10888  
QY 1103 ----- 1103  
Db 10889 GAGAAAGACTGACTCAGTGGGGCTTGGGGCTTCAAGTAGAGACTGCCAGCCCTT 10948  
QY 1103 ----- 1103  
Db 10949 CTATCTGTGCTTCCGGTCCCGACCCGACTCACCTGTCTCAGCATGGGGCAGTGGG 11008  
QY 1103 ----- 1103  
Db 11009 TGTATGTCTAGTGTAAAGTGTGCTCAGCTGATTTCTGGGGTGGGGCCCCCAGGCT 11068  
QY 1103 ----- 1103  
Db 11069 ACCCAGAGCCCAAGAGATTGGGGGAGAGATGCTTTGTGTGATGTTGCCATGGG 11128  
QY 1103 ----- 1103  
Db 11129 GATAGAGGTCTGGCGGCCAAGCAGCATGGGGCTCAGTGGGAGGGGAGCTAGGTGGCC 11188  
QY 1103 ----- 1103  
Db 11189 CACTGACCAATTACCACTTCCCCCAGCCCGCAATCTCTGTGTGTGGGGAG 11248  
QY 1103 ----- 1103  
Db 11249 TGGGAAACAGCAGTGTGGATTAATGTGTGTGTAGGTAGAGGCTGAGGGGCA 11308  
QY 1103 ----- 1103  
Db 11309 GGGGATGATTTTGTGGGTCTCAGGATGTCTGGGCCCTGTAGAAAAGTATGAAAAGA 11368  
QY 1103 ----- 1103  
Db 11369 CAACATAGGTTCCCGCGAGCAGACAGACTTGGGGGAAAGGAGCAGAGCGGTGC 11428  
QY 1103 ----- 1103  
Db 11429 TTGGGAGGAAAGGTGCAAAACATTAAGAGTGTGCCGTGATCAAGCAGAAAGTT 11488  
QY 1103 ----- 1103  
Db 11489 TTGGAAAGATGCAAGAGACTCAATCTCATCCCTGAGCCCTGAAAGAGTAGATTG 11548  
QY 1103 ----- 1103  
Db 11549 GCCCAGCTGACATACCTGTGCTCAGCACCTTATGCCAGTTGAAAGAGAGCTTTCT 11608  
QY 1103 ----- 1103  
Db 11609 TGTGTGTGTGCAAGAAATATGCAAGGCTGAATTTGACGCCCGAGGCTGGTCTGACTGTG 11668  
QY 1104 ----- 1119  
Db 1104 -----AspGlyValHisSerValIlnrAlaGlnCysAlaLeuArgValIlnrIle 1119

Db	11669	TCCCTCTCCACAGACGCGCTTACACACGCGACCGCCAGTGGCGGTGGTGTGACATC	11728	Db	12749	TGGCATGAGAGTGTGGCTTTGATTAAAGGAAACCTGAGATTGGGTGACTTGTCTGT	12808
Qy	1120	ethrAspGluMetLeuThrHisSer   ethrLeuArgGluLysMetSerProGlu	1139	Qy	1319	-----	1319
Db	11729	ATCACCGATGAGATGCTCACCCACAGCATACGCTGCGCTGTGAGGACATGTACCCAG	11788	Db	12809	GTCTCATCACACCGGTTTAGAGGTGAGGCTGGGCTTAGAACCTGTCTCTCCCACTG	12868
Qy	1140	ArgPheLeuSerProLeuLeuGlyLeuPhe   eglnAlaValAlaAlaThrLeuAlaThr	1159	Qy	1319	-----	1319
Db	11789	CCCTTCTGTGTACACACTGTCTAGGCTCTTCAATCCAGGCGGTGGCCGACAGCTGGCCAG	11848	Db	12869	AGCGAGTTCCTTTTCTGAGCTGACAGAGTACTTTTCTCCCAAGCATTAAGGGGAGAG	12928
Qy	1160	ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis	1179	Qy	1319	-----	1319
Db	11849	CCACCGACACAGGTGGTGGTCTTCAACGTACAGCGGAGACACGACGCCCCCGGGGCCAC	11908	Db	12929	ACAGAGAGCTGGGGGAACTGAGGGGGTTGAGTTCTGAGATGGGGTCTCCAGAAAAC	12988
Qy	1180	LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhe	1199	Qy	1319	-----	1319
Db	11909	ATCTTCAAGGTGAGGCTGTGCGTGGGCCACGCGCCAGGGCCCGGGGGCGGCGCTTC	11968	Db	12989	ATTGACCTGATATGGGCTGTGGGAGAAATGGCTCCGCTCCGCCACGTGCTGTAAAGA	13048
Qy	1200	LeuProSerGluAspLeuGlnGlyArgLeuTyrlLeuAsnArgSerLeuLeuThrAlaIle	1219	Qy	1319	-----	1319
Db	11969	CTGCCCTCTGAGAGACTGTGACGAGGCGCTTATCTCAACCGACGCTGTGACGGCCATC	12028	Db	13049	GGATCTGTGCTAGCGGCTTCTGTGTGGCTGTACGCCAAGTCTTTACACATCTGTCT	13108
Qy	1220	SerAlaGlnArgValLeuProPheAspAspAsn   eCysLeuArgGluProCysGluAsn	1239	Qy	1319	-----	1319
Db	12029	TGGGACACGCGCGGTGTGCTTCAACGACATCTGTCTGGAGCGCTGCGACAGAC	12088	Db	13109	CCCTGGAGTGTCACTGTCTGTCTTATGTCTGTACCCCTCTCCAGGCTCTGTGT	13168
Qy	1240	TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPhe   eAlaSerSer	1259	Qy	1319	-----	1319
Db	12089	TACATGGCGGTGTGTGCGTGTGCTTGCATCTCTCCGCCCTTCAATGCCCTCTCTCC	12148	Db	13169	CTTCTCAGAGACACCCCTGTCTCCCTTCCCTGTGGGCTTCTTCTTCAATGCCAGGC	13228
Qy	1260	SerValLeuPheArgPro   eHisProValGlyGlyLeuArgCysArgCysProProGly	1279	Qy	1319	-----	1319
Db	12149	TCCGGCTCTTCCCGGCCATCCACCCCGTGGAGGGCTGCGCTGCGCTGCCCGCCG	12208	Db	13229	CCTGTCTCTTGTCTTCCAGGGCCATGCCCACTGTGCTGTGCCAGAGCATAGGAGGCC	13288
Qy	1280	PheThrGlyAspTyrCysGluThrGlnValAspLeuCysTyrSerArgProCysGlyPro	1299	Qy	1319	-----	1319
Db	12209	TTTACCGGTGATCTTGTGACAGACCGAGGTGACCTGTCTACTCGGGGCGCTGTGGCCCC	12268	Db	13289	CACTCTTCCCGCTGTCACTTCTCTGCCAGAACTTGGCCAGACTTGTACTTCTGTGTG	13348
Qy	1300	HisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAsp   YrThr	1319	Qy	1319	-----	1319
Db	12269	CACGGGCGCTGCGCGACGCGGAGGGCGGCTACACTGCTGTGTGTGTGTGTGTGTGT	12328	Db	13349	CTGTAGGGGTGGGGGAGAGGGAGAGGAGCGGGCTTGGGGGAGGGGAGAGAGTCTTTGT	13408
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12329	GGTGGACCAAGGAGGAGGACTCATGTGGCCAGCCCTGGAAGGCTGACTGTGTGTGTGT	12388	Db	13409	CAGAGATGTGTGGGGGGGTTTGGAGGGGGGGGTTAGACAGGCGAACAAGTTTCC	13468
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12389	ACAAATCAGACAAATGCTGGCGGCTGCTCATTTCTTCCGAGTGTGTGTGTGTGTGT	12448	Db	13469	ATGAAACAAAGCCAGTCACTAGGCAACAGAGCTGAAACTATCTGTAGAACTGTGGG	13528
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12449	CTGAGAGGTCAATAATCCGCGCTGCTCCCTGTACAGACCCCACTGAGGAAGCACTGTGA	12508	Db	13529	ATGGGAGGGAGAGAAAGGTGTGATTTCCATCTGATGTGTGAGTTTCAAGTCTCTGT	13588
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12509	GGAGTGTCTTCCCTGT	12568	Db	13589	TGAATCTTGTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	13648
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12569	CTCCAGAGCCAGTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	12628	Db	13649	AGTCTTGT	13708
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12629	TGAACAGTAACTTAACCTTTCTGCGCTCGGTATTTGAATGTGAGATTAATGACCTTAC	12688	Db	13709	TCCCAAGCTTCTCTGTATCTTTCATTAACCAAGCAATCTAATGTGAGAGCTGTGT	13768
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12689	TTAGGTCAATAGT	12748	Db	13769	CCATCCCAGAGAGGAGAACTGATCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGT	13828
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	13829	GGAGGAGAGAGAGGCTCCCAAGCCCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGT	13888	Db	13829	GGAGGAGAGAGAGGCTCCCAAGCCCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGT	13888

QY 1319 ----- 1319  
 Db 13889 ATTATCCCTTGGGGGCCAGACACCCCAAGTGTCTTGGTATGTATCATTCGAAGC 13948  
 QY 1319 ----- 1319  
 Db 13949 TCCCGGATCCGGGCTTACAACCAAGTGTCCAGCAGGCTGTGCTTATGAACCCAAA 14008  
 QY 1319 ----- 1319  
 Db 14009 TGGGTAAGGAAAGTGTGGGGACTCCGAGAACTTTCAGGGAGCAGTGCCTGAGCTG 14068  
 QY 1319 ----- 1319  
 Db 14069 TGGCGAGAGCGGTCAGTAGCCAAACAGGGTCCATGAAAGGCTGGGACCATAGCTC 14128  
 QY 1319 ----- 1319  
 Db 14129 TCAGAGAGATGACACAGAGAGAGTGGGATGCTGGAGTTTGGGAGTATTACG 14188  
 QY 1319 ----- 1319  
 Db 14189 AGGGTCAAGGAGGCCATGAGAGCTATCACTGGGTTGTGGGGGTCTCACTTCCCTCT 14248  
 QY 1320 ----- 1325  
 Db 14249 GTCTCAGGTACTTGTGTACTCACTGCTTCTCTTCTTCAGGTGAGCACTGTAGAGTG 14308  
 QY 1326 SerAlaArgSerGlyArgCysThrProGlyValCysLysAsnGlyValThrCysValAla 1345  
 Db 14309 AGTGGCTCGCTCAGAGCGCGTTGCACCCCGGGTGTCTGCAGAAATGGGGGCACTGTGTCAAC 14368  
 QY 1346 LeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPheGlyLysProThrCys 1365  
 Db 14369 CTGCTGTGGGGGTTCAAGTGCATGCCCATCTCGAGACTTCAGAGAGCCCTATCTGC 14428  
 QY 1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGln 1385  
 Db 14429 CAGGTACACCGCGAGCTTCCCGGCCACTCTTCATCACTTTGGCGGCTGGCGAG 14488  
 QY 1386 ArgPheHisPheThrIleuAlaLeuSer----- 1394  
 Db 14489 CGTTTCACTTCACTCCCTGCCCCCTC-GTGAAGTGCCTGGCACTGGGGGTGGAGTGGG 14547  
 QY 1394 ----- 1394  
 Db 14548 CCTGGTGGGATCTGAGGTGCTCTGTCTGTGTGAAGTAAAGTCAAGCACTGAA 14607  
 QY 1394 ----- 1394  
 Db 14608 AGCGTGAAGAGGCTGCAGGAGAGCCAGAGTCTGCTTGAAGTTGCTGTGTCTGCGC 14667  
 QY 1394 ----- 1394  
 Db 14668 AGCAGATCTCCCTCCCTGCGCTGTAGGTGCGGGGTCTGCATCCAGTGGGTAC 14727  
 QY 1395 ----- 1402  
 Db 14728 CCCATTCCTGCCCCCATCCCACTCTGTTCAGGTTTGCCCAAGAGAGCGGCAAGG 14787  
 QY 1403 LeuLeuLeuTyrAsnGlyArgPheAsnGlyLysHisAspPheValAlaLeuGlyValIle 1422  
 Db 14788 TTGCTGTGTATCATATGGGCGTTTCATATGAGAGCATGACTTGTGTGCTCGAGGTATC 14847  
 QY 1423 GlnGluGlnValGlnLeuThrPheSerAla----- 1432  
 Db 14848 CAGAGCAGAGTCCAGCTCACTTCTCTGC-AGGTATCAAGTTCCTCCCATCTCTGCGC 14906  
 QY 1432 ----- 1432  
 Db 14907 CATCTTCAAGAGCCCAAGTCTTCAAGCCCTGACCCCAAGGACATATCTATCAGCC 14966

QY 1432 ----- 1432  
 Db 14967 AAATGTGGGCCAAGCCCAAGCACTGGGACCCCAAACTTCCTCATCTTACCCCTGCA 15026  
 QY 1433 ----- 1442  
 Db 15027 GGCAATTCAGCTCACTGTGCTCTTCCCAAGGGAGTCAACACAGGCTGTCCCATTC 15086  
 QY 1443 ValProGlyGlyValLysArgGlyValIleThrPheValGlnLysValLysValLys 1462  
 Db 15087 GTGCCGAGAGAGTCAAGTATGACCAATGATCAGTGCAGCTGAATACTACATTAAG 15146  
 QY 1462 ----- 1462  
 Db 15147 GTGGTGTGAGAGGCAAGAGGTTGGGGGTCTGTCTTGTCCCTCAGTGTCTTACCCAGC 15206  
 QY 1462 ----- 1462  
 Db 15207 CCTGAGTGGCAATTCCTCCAGGCTTGGGTGGCTGTGCAGGCAATTCGTGAGAGA 15266  
 QY 1462 ----- 1462  
 Db 15267 AGGAGCGGCTGGAGAGGTGCCACTTGTCTGGGCAATTCACCCCACTGGGCACTGCTG 15326  
 QY 1462 ----- 1462  
 Db 15327 ATGCTCTGTTCACCTAGGAGCGGCAAACTTGGACACTGAGGAAGTATTACATGCA 15386  
 QY 1462 ----- 1462  
 Db 15387 TAAACAGTGAGCAGACGTGTGTATATGATGTGTATCGTGTGTGGGCAATGTGG 15446  
 QY 1462 ----- 1462  
 Db 15447 GTGCACACAGAGCAACCAAGTGTGCTGTGCGCATGAACCTATGGGCTCTGTACTCACTG 15506  
 QY 1463 ----- 1472  
 Db 15507 TCCCCCTTCTGTGCTCTTCTCTGTGCACAGCACTGTGGGTGAGAGCGGCTCCACAG 15566  
 QY 1473 GlyProSerGlyLysValLysValAlaValThrValAspGlyCysAspThrGlyValAla 1492  
 Db 15567 GGGCCATTCAGAGCAAGAGTGGCTGTGTACCGTGAATGCTGTGACACAGAGTGGCC 15626  
 QY 1493 LeuArgPheGlySerValLysGlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGly 1512  
 Db 15627 TTGGGCTTCGGAATCTGTCTGTGGCACTATCTCTGTGCTGCCAGGGCAACCAAGGTGGC 15686  
 QY 1513 SerLysLys----- 1515  
 Db 15687 AGCAAGAAGTGAGCAGGGGAAAGGCGCAGGATGGGTGAAGTGAAGGCAAGGCCAGGCA 15746  
 QY 1515 ----- 1515  
 Db 15747 GGAAGGAGGTGAGAGAGGTGGGCTTTTGGGTTCTCTCTCTTGGCTCTTAAAGCC 15806  
 QY 1515 ----- 1515  
 Db 15807 CTGCTCAGCCCTGGGGATGGGGTCAAGACTGGGGAATCAGAGCAAGGCCCTGTGGGAG 15866  
 QY 1515 ----- 1515  
 Db 15867 TCTGGAGAGGGGCTGAGTGTTCAGAAAGAACAGTTCAGGTAGGGGTGAGCCCTTCT 15926  
 QY 1515 ----- 1515  
 Db 15927 GAGCACTGCTTGCAAGGCTCTTGGACACTGTGCTACGCCGTTCGTGTCTGTGGCTGG 15986  
 QY 1515 ----- 1515  
 Db 15987 CGGAGAGCAGAGGAGGACCAAGAGCTCCAGGCTCCCTTGTGCGGCACATCTTGG 16046  
 QY 1516 ----- 1532  
 Db 1516 -----SerLeuAspLeuThrGlyProLeuLeuLeuGlyValProAspLeuPr 1532



```

Db 16047 CTCCTCCAGGCTCTGATCTAAGCGGCGCTCTGCTACTAGCGGGGCTCGACCTGCGC 16106
QY 1532 OGluSerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAs 1552
Db 16107 CGAAGAGCTTCCAGTCCGAAATCCGGAGTTCCGTGGCTGCAATCGGAACCTCAGAGTGA 16166
QY 1552 pSerArgHisIleAspMetAlaAspPheIleAlaAsnAngIyThrVal 1568
Db 16167 CAGCCGGCACAATGACATGGCTGACTTCATTCCTCAACAAATGACACCGTGCCTGTATGGG 16226
QY 1568 1568
Db 16227 GGCCCGGGGTGAGCCAGGCTGGATCCAGTGTGAAGAGAGGGGCTGTGCTGTGTG 16286
QY 1568 1568
Db 16287 AGGATGCCAGATGAGGGTCTCAGAGGCTCTCTAATCTGCTCCACAGGCTTTCTTT 16346
QY 1569 1575
Db 16347 AGAGCCCAAGCTCAGGTCCAGTCAAGCTGCTTCTCTCCATTCAGAGCTGCTGCGCAAGA 16406
QY 1575 1595
Db 16407 YSAAsnValCysAspSerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTyrAspA 1595
QY 1595 1646
Db 16407 AGAAGCTGTGTACAGCAACACTTGCACAAATGGGGGCACTTGGCTGTGAACAGTGGAGC 16466
QY 1595 1611
Db 16467 lAphSerCysGluCysProLeuGlyPheGlyGlyCysSerCysAlaGln 1611
QY 1611 1656
Db 1611 CGTTAGCTGCGAAGTGTCCCTCGGCTTTGGGGGCAAGCTGCCCAAGGGTAGAGAGG 16566
QY 1611 1611
Db 16527 GCGGCTGTAGAGGCCAGACAGCTGGGGTGCATCAACATCTGGAACTGGCAGGGTTGGG 16586
QY 1611 1611
Db 16587 GCAGGCACTGGCAGGGCTCTGTGAGCGGGCTGTGGGTGAAAAAGCTGTCTGGGCGAG 16646
QY 1611 1611
Db 16647 AGCCCTGAAGAGGCAACAGAGAACAAAGGGCCAGTCAAGAGCGAGGCTGGGCAAGC 16706
QY 1611 1611
Db 16707 GTGGGGCGGTCCAGGCGAGTACGCACTTTGAGGGCGGGGCTGATGAGGGAGTGGGC 16766
QY 1611 1611
Db 16767 TCTGCTCTCCGTGAGCGCGCAACGCTGAGCATCAAGCCCCAGGGCCCTTGAGGCTCC 16826
QY 1612 1620
Db 16827 ACCCTCAACGTGCTTTCCTGCTCCCTCCAGAAATGGCCAAATCCACAGACATTTCTG 16886
QY 1621 1640
Db 16827 GlysSerLeuValAlaTyrHisGlyLeuSerLeuProIleSerGlnProTyrTrpLeu 1640
QY 1687 16946
Db 1687 GGCAGACACCTGTGGCTGTGGCATGGCTCTGCTGCCATCTCCAAACCTGTGTAAGCTC 16946
QY 1694 1660
Db 1694 SerLeuMetPheAspThrArgGlnAlaAspGlyValIleLeuGlnAlaIleThrArgGly 1660
QY 1661 1666
Db 1661 ArgSerThrIleThrLeu 1666
QY 17007 CGACACACATCAACCTCAAGGTATGCAATGGAAGGGGCGTGGCCCTGGGCTGCGCATTA 17066
QY 1666 1666
Db 17067 GGGCCCTGTAGCTCTAGAGCGGCTGGAAGAAATGTGGCTGGCAGAGTCTGGGCGAGGGGT 17126
QY 1666 1666
Db 1666 1666

```

---

```

Db 17127 GGGACATATAGAGCGCGCTGGAATCCGTTGGAAAGTCAATGTGCCACTGTGGGCTT 17186
QY 1666 1666
Db 17187 GAGGAAATTAATACGCTTTCCTCTATAGGCTCAGCTGGGCAAGGGATGGTTGAT 17246
QY 1666 1666
Db 17247 GGTCCAGAAAGGGAAGCTGTAGCTGAGAAACAGTATTAAGCCGCTGTGACCTGT 17306
QY 1667 1680
Db 17307 TTTCTTCTCTGCTGTGCTCACTAGAGAGGCCACGTGAAGCTGAGGTGAAGGCA 17366
QY 1681 1700
Db 17367 GGGCTTCAAGGCTCTCTCTCTCCCTGTGAGGCGAGGCGGCGCAATGACGTGATGGGAC 17426
QY 1701 1720
Db 17427 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720
QY 1742 17486
Db 17427 CATGCAAGCTGGCACTGGAGGCGAGCGGGGCGCGGCAATGCAATCTGTCTTCAT 17486
QY 1721 1740
Db 17487 TyrGlyGlnGlnArgAlaGlyGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740
QY 1741 17546
Db 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAspGlyPheArgGly 17546
QY 17547 AACATTAACGTGGCGGAATTAACCTGGGCAAGCGGCGGTGTGGCTTGCGGGG 17606
QY 1761 1761
Db 17607 TGTGGAGGTAGGTCTGCTGCTGCTCCATCCCTCCCAACCACTGAGCGCC 17666
QY 1762 1771
Db 17667 GCTCCACTGAGGGCAACAGCTGCTGTCTCCCTCCAGAGGTGTGGGTGAGCGATAGGCC 17726
QY 1771 1791
Db 17727 OGluGlyValAsnSerLeuAspProSerHisGlyGlySerIleAsnValGluGlnGlyC 1791
QY 1778 1786
Db 1778 TAGCCTGCTGACCTTGTGATCTCAACCGGTGCTGTAAACAGTATTGCAACAAGA 1786
QY 1811 1820
Db 1811 pTyrAspSerTyrSerCysSerCysAsp 1820
QY 17847 CTGGACAGCTAATCTGTGACGTGTGA-TCAGGTATGTAAAGATCAAGGCAACGGGC 17905
QY 1820 1820
Db 17906 AGTTATCAGGTGCTGGGGCCACATGCTGGCTGTGTGTACCTGGGGCATGGGGCA 17965
QY 1820 1820
Db 17966 TCACACACCTCTGTAGTCTTAGTTGCTTACACAACAAAGCATAGCAGTGAATAGA 18025
QY 1820 1820
Db 18026 TGTATGTGAAAGAAACAGATGAAGGTTGATGAAAGTGTGCCAGGGAGGGTGGG 18085
QY 1821 1831
Db 1806 GGAATGAGCCTCTGTGCTCTGTGCTGCTCCAGGTACTATGTGTAACAATCTACTATGT 18145
QY 1831 1851
Db 18146 lCysAspLeuAsnProCysGlyHisGlnSerValCysThrArgIleProSerAlaProH 1851
QY 1851 1868
Db 1851 ArgIleThrCysGluCysProProAsnIleLeuGlyProTyrCysGlyHis 1868
QY 18206 TGGCTAATCTGGAAGTGTCCCAATTAACCTTGGGCGCAATCTGTGAGACAGGTAAAC 18265

```



Db 20426 AGTAGGGGAGTGTGCTGCTGAGCCAGAGACACCCAGAGATGGGGGCTGGCGCTGT 20485  
 QY 2004 ----- 2004  
 Db 20486 CCTGAGCTGGCAATGAGCAGGTTCAATAGGAGAGAAATGGCGGAGAGACAAAGTGTGGC 20545  
 QY 2004 ----- 2004  
 Db 20546 TGGAGACAGGAGAGATGGGGGCTCAAGAGAGTCGTATGCCCTGTTCCCTTAGAGCCCAATG 20605  
 QY 2004 ----- 2004  
 Db 20606 TGTGAGCTGATCTTGGGAGAAATTTTGGCCCTGGTGTCTACATACTGTCTTCTGTTT 20665  
 QY 2004 ----- 2004  
 Db 20666 CACAGAGGCGCTTGTGTTGTAGTTACTGTCCCGGCCCATGAGGTGGCTCCAGCAG 20725  
 QY 2004 ----- 2004  
 Db 20726 AACCTCTCAGCCAGTCTCAGGGGCTTCTGTGTCTCTCCCTGAGGCGGCGCTTTTGCT 20785  
 QY 2005 ----- 2005  
 Db 20786 TCTTCCCGCCAGAGACTGTGTGGCGCACTGTATGAGCAGAGGGGGTGGCTCCCGCAA 20845  
 QY 2021 snlLeuPheasnCyethrSerIleThrPheSerGluLeuLysGlyPhe----- 2036  
 Db 20846 ACCTTTCACTGACAGTCCTCAACCTTCTCAGAACTGAGAGGGCTTCTGTAAGTGAACCC 20905  
 QY 2036 ----- 2036  
 Db 20906 CTGATCTCCATCTTTCCTGTCCTTCCGTCGTAAGTCTCACTTGGCCCACTCCATCTT 20965  
 QY 2036 ----- 2036  
 Db 20966 TGAAGACGGGCTTCTGGAATTCAGCCTGTGTGTCTGGGCTCCAGCTGGAGAGGCC 21025  
 QY 2036 ----- 2036  
 Db 21026 GTCTCTCACTCTGAGGTCCTTTTGTCTCAGAGTTCCTGCCCTCACTCTGCTCTCCG 21085  
 QY 2037 ----- 2037  
 Db 21086 CTCCGTCCTGTCTAGGCTGAGCGGCTACAGCGGAATGAGTCAAGGCTTCAAGGCGGC 21145  
 QY 2052 SerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnIleThrAlaGlyTyrPheGly 2071  
 Db 21146 TCCGAGCAGCTAGCCCTGCTCTCTGCGCAACGCCAGCAGCAGCAGCAGCTGACTTTCGCGC 21205  
 QY 2072 SerAspValIleValAlaTyrGlnIleLeuAlaThrArgLeuLeuAlaHisGlnSerThrGln 2091  
 Db 21206 AGCGAGCTGACAGGTGGCTTACCAAGCTGGCCAGCGGCTGTGGCCAGAGAGCACCGAG 21265  
 QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGln----- 2106  
 Db 21266 CGGGGCTTGGGCTGTCTGCGCACAGAGACGTCACTTCACTGAGGTGGGGCTTGGAGGA 21325  
 QY 2106 ----- 2106  
 Db 21326 TGCAGGGCTGGCTGTAGATAGGGGTCATGTGTAGTAGAGCTGTCTATGCGAATGGGG 21385  
 QY 2106 ----- 2106  
 Db 21386 GGGCAGAGGGGGCGCTCCCATCCACCTAACAAGAGGTCCTCGTGGAGAGAGGCTTCCA 21445  
 QY 2106 ----- 2106  
 Db 21446 TGAACCTGTGACCCCTGGGCCAGACCTCTTCCCAACCTTCTCATATGATATATGTG 21505  
 QY 2106 ----- 2106

Db 21506 TGTGGGCAATGTCTGACCCAGAGCAGAGGCTTGCCTGTGGTACCATAGGCTGGAGGA 21565  
 QY 2106 ----- 2106  
 Db 21566 AGCAGTATCTCAGACATGTGTGTGGTCCAGGACAGGGCTGGAGACTTTATATAGA 21625  
 QY 2106 ----- 2106  
 Db 21626 GAATGAGAGAGGGCTTAGGGGCGAGAGCATTCAGTGACTGAGTGGGCAGTGGCTCA 21685  
 QY 2106 ----- 2106  
 Db 21686 GAGCATTTTCAAGGGAGCAACAGTGAACAGAGGAGGGGCTGTGGGAGCAGGGA 21745  
 QY 2106 ----- 2106  
 Db 21746 CTGGCGCGGCGACAGGCGGGGCTTCATGTGGGAGTTCAGGTTGGGTTGTGACCA 21805  
 QY 2107 ----- 2107  
 Db 21806 CCCACCGTAGCCTTGCCCAACCCAGAAATCTGCTGGGGTGGCAGCGCCTCTGGACAC 21865  
 QY 2118 rAlaAsnLysArgHisTyrGlyLeuIleGlnGlnThrGlyLeuIleTyrAlaIlePheLeu 2138  
 Db 21866 AGCCAAACAAGGGCACTGGGAGCTGATTCAGACAGAGGGGTGGCAACCGCTGGCTGCT 21925  
 QY 2138 uGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSe 2158  
 Db 21926 CCAAGCATATAGAGCTTACGCCAGTCCCTGGCCCAAGAACATGGGGCACACTTACTTAG 21985  
 QY 2158 rProPheThrIleValIleThrProAsnIle----- 2167  
 Db 21986 CCCCTTCACTATCTGACGCGCCAACTGTGAAGCTGTGTCCTGGTTGGAGGGGTT 22045  
 QY 2167 ----- 2167  
 Db 22046 TGTGAGGAGTCCCGCAAGAGCGGCTGTGCTGGGTCCTGCTCAGAGTCCG 22105  
 QY 2168 ----- 2168  
 Db 22106 ATCTGTGACATCCCTTCTTACTTCACTTCCGTAAGTGCCTTGGACAAAGGAACCTTG 22165  
 QY 2180 lAgIyAlaLysLeuProArgTyrGlyAlaLeuArgIlyGlnIleProProAspLeuGlnT 2200  
 Db 22166 CTGGGGCCAGCTGCCCGCTACGAGGCGCTGCGTGGGAGCAGCCCGGACCTTGAGA 22225  
 QY 2200 hTrhValIleLeuProGlnSerValPheArg----- 2210  
 Db 22226 CAACAGTCATTTCTGCTGAGTCTGTCTTCAG-AGTCAGTGGTGGCCATGGAATTGAGTTG 22284  
 QY 2210 ----- 2210  
 Db 22285 GGAAGTCGACCCCAAGTGTCTGTGACACTCACAAGAGCAGGGGCCAGCTAAGTGCAC 22344  
 QY 2211 ----- 2211  
 Db 22345 AGTGTCCCTTCCAGAGACCGCCCGTGTGCAAGGCCGCGCAGGCCCGCGAGAGGCCAGG 22404  
 QY 2226 lueProGlnGlnLeuAlaArgArgIleArgArgHisPProGlnLeuSerGlnIlyGlnAlaV 2246  
 Db 22405 AGCCAGAGAGAGCTGACCGGAGCAGAGCGGCAACCGAGCTGAGCCAGAGGTAGAGCTG 22464  
 QY 2246 aAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProA 2266  
 Db 22465 TGGCCAGGTCAATCATTTACCGGACCCCTGGCGGGCTACTGCTCATATACATACCTTG 22524  
 QY 2266 ePlyAspSerLeu----- 2270  
 Db 22525 ACAAGCGCAGCTTGAAGTCAAGACTAGGGGACAGGTGTGGTATAGGGTATGGGTGGGC 22584  
 QY 2270 ----- 2270  
 Db 22585 GGTGAGTCTGAGGCAATGAGGGGGGTGGGGGCGTCTTCCCAAGTCATGTGACTGCCGTGG 22644

QY	2271	-----ArgValProIyAaRgProIleIleAaThrPro	2281
Db	22645	TGACTGTGCACCTGACTGACCTGCCCAATCAAGATGTCCTCCAAAGCCGATCATCAACAACCC	22704
QY	2282	ValValSerIleSerValHisAaPAspGIuGIuLeuLeuProAaRgAlaLeuAaPlyPro	2301
Db	22705	GTGGTGAAGCATCAACGCTGCATGATGATGAAGAGCTTGTGCCCGGGCCCTGGACAACCC	22766
QY	2302	ValThrValGlnPheAaGLeuLeuGluThrGluGluAaThrIyPheIyPheValPhe	2321
Db	22765	GTCAACGGGTGCAGTTTCCGCTGTCTGGAGACAAGAGAGGAGCAAGGCCACTGTGTCTTC	22824
QY	2322	TrpAaHisSerIleLeu-----	2327
Db	22825	TGGAAACATTCAATCTGTGTAGACTGCACGTGCCCTCGCCCCCTCAGGCTTCGGGCTGA	22886
QY	2327	-----	2327
Db	22885	GTCCAGGCCCCGTGATGCTCAACCTCCTTGTCCTCCGTACCCGTGGCTTCTCAACAAG	22944
QY	2328	ValSerGIyThrArgIyLTPsSerAlaArgGIyCyseGIuValValPheAaAaGluSer	2347
Db	22945	GTCAATGTGCACAGGTGTGCTGTGTGGCCAGAGGCTGTGAATGCTCTTCGGCATGTGAAC	23004
QY	2348	HisValSerCyseGIuCyAaHisIleMetThrSerPheAlaValIleuMetAaPValSerArg	2367
Db	23005	CACCTCAAGCTGCCAGTGCACACACATACAGACGTTGGCTGTCTCATGAGACGTTTCTCG	23066
QY	2368	ArgGIu-----	2369
Db	23065	CGGAGAGTCGGAGCCCAAGAGGGCAGCTGCAGAGCCGTGGGTGGACCCAGGGCAGGGG	23124
QY	2370	-----As	2370
Db	23125	CTGGGTGCTCAGGTCTGACCTTCTTAATCCCTGAGCCCTGTGCACCTACTCTGCAGAA	23184
QY	2370	ngIyGIuIleuProLeuIySthIeuthrYrValAlaLeuGIyValThrLeuAlaI	2390
Db	23185	TGGGAGATCCTTGCACATGAAAGACATGACATAGTGAGCTCTAAGTGTCACTGTGCTGC	23244
QY	2390	AlaLeuLeuThrPhePhePheLeuThrIleuAaArgIleuAaArgSerAaGlnHisGI	2410
Db	23245	CTTCTGTCTCACTTCTCTTCTTCTTCTTGCACCTCTTGCATCTGCGCTCCCAACACACCG	23304
QY	2410	YIleArgArgAaMetuThrAlaIaLeuGIyLeuAlaGlnLeuValPheLeuLeuGIyIle	2430
Db	23305	CATCCGACGTAACTGTACAGCTGTGCCCTGGGCTGGCTCACTGGTCTTCTCTTGGGAAT	23366
QY	2430	EAaGlnAlaAaPLeuPro-----	2436
Db	23365	CAACCAAGGCTGACCTCCTGTAAAGTGTCTACTGCCCCAGAAATGTCCCACTTCTTC	23424
QY	2436	-----	2436
Db	23425	AGGCGGCTCCCAAGCCCCCACTGGCAACCCCTGTCTGTGCACGATGAATCTTAATAG	23484
QY	2436	-----	2436
Db	23485	TGCGTAGTGACGACCTGGGCCCAAGGATTTCTCTTCTGTGTGGTCCCCCGGAATCCCCAG	23544
QY	2436	-----	2436
Db	23545	CACCTGCTTGGGCCCAAGCTTCCCTGGAAGCAGATTCCAAACACCAAGGCTCTTCAT	23604
QY	2436	-----	2436
Db	23605	GCCTGACCCGAAGCAGAGCTGTCTGTGGCGGGCCCCGGTGGCTGACTGTGCCCTGGGC	23664
QY	2437	-----PheAlaCyseThrValIleAlaIleLeuLeuHisPheLeuIyYrLeuCyseT	2453
Db	23665	TGGGCGCTCAAGTTGCTGCGACATGTCAATGCAATCCGTCTGCACCTTCTGTATCTCTGCA	23724

QY	2453	hrpHseSerTPAlaLeuLeuGluabAleuthiStLeuYrAaGAlaLeuThgluValArga	2473
Db	23725	CCCTTCTCTGGGCTCTCTGAGGCTTGGACCTGTAACCGAGGACATCACTGAAGGTGGCG	2378
QY	2473	spValaenThglYProMeArGPhetYrTYrmeLeuGlYTPGlyValProAlahei	2493
Db	23785	ATGTCACACCGGCGCCATGGCTTCTACTACATGCTGGGCTGGGGGTGGCTGGCTTCA	2384
QY	2493	leThr-----	2494
Db	23845	TCACAGGTACTCCACCCATTCACAGTCTTGGGGTCCACATGCCCTGGGTCCACTTGT	2390
QY	2494	-----	2494
Db	23905	GCATGTTCTTCCACCCACATACAGGCGCTGAGGCCCAATCCCATGCCCCAGGCG	2396
QY	2494	-----	2494
Db	23965	CCTTATTCCACAGTGTCCCTCTGGTTTAAACCAGACTCTGACCGGCCAACCCAGGCTCA	2402
QY	2494	-----	2494
Db	24025	CTCCCTTATCTGGGGAGATGGGTAGGGGCGGATGGTGGCAGAACTTTTTCATGCTT	2408
QY	2494	-----	2494
Db	24085	CATGCCGTGGCCGTGTGAGGCCAAGCTGGCCGAGGCCACTTCCAGGCGCCCTCATCACCC	2414
QY	2495	-----GlyLeuAlaValGlyLeuAbpProGluG	2504
Db	24145	CTCCACTGTCCCGCTGTGTCTCCATGCTCCAGAGGCTA-GCGGTGGGCTGTGACCCGAGG	2420
QY	2504	LYTYrGlyAenProAbpPheCystrProLeuSerIleTyraAbpTrLeuIleTrioSerPhea	2524
Db	24204	GCTACGGGAACCTGACATCTTGTGCGGCTTCTTCATCTATGACAGGCTCATCTGGAATTGG	2426
QY	2524	IaGlyProValAlaPheAlaValSer-----	2532
Db	24264	CTGGCCGGGTGGCTTGTACCGTCTCGGTGAGTGAAGTACAGGTGGGTGGTGTACCTGT	2432
QY	2532	-----	2532
Db	24324	GGCCCTCTTGTCTGTCTCTCTGCTGCTGGGCGCGGTGTCTCAGGACTCCCTTAAAGA	2438
QY	2532	-----	2532
Db	24384	ACAGCTGAGGCCACGGGCGCTGGGTGGGTAAGTGTTTGTGCCCATGAGACCC	2444
QY	2533	-----MetSerValPheLeuTYrIleLeuAlaAlaArgAlaS	2545
Db	24444	CCATGGCCCTMGCCCTGCTAGTGAATGTCTTCTCTGTAATCTTGGCGGCGCGGCT	2450
QY	2545	erCybAlaalaGlnArxGlnGlyPheGlyIuTyrsIysGly-----	2557
Db	24504	CTGTGTGCTGCCGAGCGGCAAGGCTTTGAGAAAGAAAGTCTGTGTAGATTAAGGTTGGG	2456
QY	2557	-----	2557
Db	24564	GTGCTGGGCGCATGGAGGACACAGCATGGAGATCTCATGCGCGGACATCATGGCTGT	2462
QY	2558	-----ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaT	2575
Db	24624	CCCTATCCCAAGC-TCGGGCGTGGAGCCCTTCTTGGCGGCTCTGTCTGTCTAGAGGCCA	2468
QY	2575	hrTrioLeuLeuAlaLeuLeuSerValAaSerArpTrileuLeuPhehIeTYrLeuPheh	2595
Db	24683	CGTGGCTGTGGCACTGCTCTCTGTCAACAGGCAACCCCTCTTCCACTACCTCTTGG	2474
QY	2595	IaThrCybAenCybIle-----	2600
Db	24743	CTACTCTCAATGTACTCAGATCACTGGCCAGCGTGTGAGGAAGGAGGACCTGGGCT	2480
QY	2601	-----GlnGlyProPheIleP	2606

```

Db      24803 GTGGATGCTGAATATGACAGACCGTGGCTCTTGGCCCTGAGCGGCGCCCTTCATCT 24862
Oy      2606 heluSerTyrValValLeuSerLygIuValArgIysAlaLeuIysLeuAlaCysSerA 2626
Db      24863 TCCTCTCCTATGAGGAGCTTAGCAGAGAGAGTCCGGAAGACCTCAAGCTTGCCTGAGCC 24922
Oy      2626 rglYsPProSerProaPProAlaLeuThrIysSerThrLeuThrSer 2642
Db      24923 GGAAGCCCAAGCCCTGACCTGCTGTGACCAACCAAGTCCACCTGACCTCGGTGAGGAGAC 24982
Oy      2642 ----- 2642
Db      24983 CAGGGGTCTCAGAGAGGCGGTGAAGGAGAGAGGAGGAGGCCAGTAGAGCTGTACT 25042
Oy      2642 ----- 2642
Db      25043 TTTGGCCCACTCCCTTCTCTTTCTTCATCCCTTCTGGAAGGTGAAGGAGAGGTG 25102
Oy      2642 ----- 2642
Db      25103 AAATGCTGTGTTCCGCTGGGAGAGAGATGTTGAAAAGGTGTGTAGATTTTGACTTGA 25162
Oy      2642 ----- 2642
Db      25163 GGGAGATGGGGCTGAGACAGATATATCAAGGTAAATAGTGCCTTGACCTGAGATCTTCA 25222
Oy      2642 ----- 2642
Db      25223 TGAATTTAGTAAATTTCTGTCTCCAGAGGTCTCTAGAGAGCAATTTCTGTCTCAT 25282
Oy      2642 ----- 2642
Db      25283 CTTTTCAGAGCCATTATATAGTAGTCCAGATCCCTGGGGCATGACAGGGTAGACACA 25342
Oy      2642 ----- 2642
Db      25343 GCTGTGGCAGAGCTGTGACCACTGCCAGGCTTCATATANGGGAACAGCGGAGAGAGGCG 25402
Oy      2642 ----- 2642
Db      25403 CCTCTCTTTCAGACCTGAGCATGTGGAAGGTGAATTTAGGGGCAAGTTCCCTCAACCTC 25462
Oy      2642 ----- 2642
Db      25463 CTTCTCCAGCGTGGGGCAGACCAATCCCACTTACTGACCTCTGTCTCCCTG 25522
Oy      2643 ----- 2643
Db      25523 CCTAGTCTTACAACTGCCCCCAGCCCTTACGAGATGGGCGGTGTACCAAGCCCTTACGAG 25582
Oy      2661 spSerAlaGlySerLeuHisSerThrSerArgSerGlyIysSerGlnProSerTyrIleP 2681
Db      25583 ACTCGGCGCGGCTCTCTGACACACACCAAGTCGCTCGGGCAAGAGTCAAGCCCAAGCTACATC 25642
Oy      2681 rhpLeuLeu----- 2684
Db      25643 CTTTCTGTGCTAGGTGAATCCCGAGATGGAGGGGTGAGAGAGGGAGAGAGGGCCCAAG 25702
Oy      2684 ----- 2684
Db      25703 CATGCTGACCCAGCCAGCCAGCTGTGTGGAGTTGAGAGACACACTGTGGCTGACGT 25762
Oy      2684 ----- 2684
Db      25763 GGGGGCAGCTTGATTAGAGCTGTAAAGGACCCACAGCAGAGAAACAGATCCCAAGGGG 25822
Oy      2684 ----- 2684
Db      25823 AGAGAGAGACTGGGACCTGGGCAAGGGGCAAGGCTGACCCCTCAGAGATGTCTCATC 25882
Oy      2685 ----- 2685

```

```

Db      25883 TTCTAGGAGAGATCCCGCACTGAACCTGGCCAAAGGCGCCCTGGGCTGGGGATCCAG 25942
Oy      2703 lYSerLeuPheLeuGluGluGlnAAspGlnIleAAsp----- 2715
Db      25943 GAGGCTGTTCTTGAAGGTCAAGACAGCAGACATGG-TAGAGACAAACGCTTGGCCA 26001
Oy      2715 ----- 2715
Db      26002 CCACAGAGGCGAGTTGGCTTTTACTGAAGGTGGGTGAGAGGTGGGCTGGGCTGT 26061
Oy      2716 ----- 2716
Db      26062 GATCTCTCCCTGCGCTCTAGATCTCTGACAGGACTCCGACAGTACCTGTCTTAGAAG 26121
Oy      2728 spAAspGlnSerGlySerTyrAlaSerThrHisSerSerAAspSerGlnGluGluGluGluG 2748
Db      26122 AGACACAGATGGCTCTTATGCTTCTTACCTACTATCAACAGATGAGAGAGAGAGAGAGG 26181
Oy      2748 lUglUglUglUglUAlaAAspPProGlyGluGlnGlyTyrAAspSerLeuLeuGlyProG 2768
Db      26182 AGAGAGAGAGAGAGGCGCGCTTCCCTGGAGAGAGGCGCTGGGATAGCTGTGGGCGCTG 26241
Oy      2768 lYAlaGluArgLeuProLeuHisSerThrProLyS----- 2779
Db      26242 GAGCAGAGAGACTGCCCTTGACAGTACTCCAAAGGTGGGCGCACACTGGGCGCTGGGC 26301
Oy      2779 ----- 2779
Db      26302 CTTTGGGGCAGTGGAGAGAGAGAGTGGCGGTCTTGGGAAGCAGAGACTGGGGTGT 26361
Oy      2780 ----- 2780
Db      26362 GGCTCTGTGCTATCTGCCCTCTGTGGGCTCATCTACTTCTTCCCAACAGATGGGG 26421
Oy      2782 lYProGlyPProGlyIyValAAspProThrPProGlyAAspPheGlyThrAlaIySgluSers 2802
Db      26422 GCCAGGGGCTGGCAGAGCGCCCTGGCCAGGAGACTTTGGGACCAACAGCAAAAGAGATA 26481
Oy      2802 erGlyAAspGlyAlaPProGluGluArgLeuArgGlyAAspAlaLeuSerArgGluG 2822
Db      26482 GTGGCAAGGGGCGCCCTGAGAGAGCGCTGCGGAGATGAGATGCCCTGTCTGAGAGG 26541
Oy      2822 lYSerLeuGlyProLeuPProGlySerSerAlaGlnProHisLyS----- 2836
Db      26542 GGTCCCTTAGGGCCCTCTTCCAGGCTCTTCTGCCAGCTCAACAAGTGAAGGGGCAACC 26601
Oy      2836 ----- 2836
Db      26602 CCAAGTCCGAGCTCCCTAGTCAAGCAAGCTCATATCTCATCTCTGTGGCGGCAAC 26661
Oy      2837 ----- 2837
Db      26662 TCACAGCCCGCGCCCGGCCCAACAGCATCTTAAAGAAAGATGTCTGCCCACTACAGCG 26721
Oy      2849 lUlySerSerLeuLeuArgLeuProLeuGluGlnGlyThrGlySerSerArgGlySers 2869
Db      26722 AGAAGAGCAGCTCTCTCGGCTCCCTCTGGAGCAATCAACAGGGGTCTTCCCGGGGCTCCT 26781
Oy      2869 erAlaSerGluGlySerArgGlyIyGlyProProAAspProProAAspArgGlnSerLeuG 2889
Db      26782 CGGCTAGTAGAGGGGAGCGGGGAGGCGCCCTCTCCCGCCCAACCGCGGAGAGGCTCC 26841
Oy      2889 lUglUglUglUglUValIleCProIleAlaIleCserIleValAlaGlyThrValAAspG 2909
Db      26842 AGAGCAGCTGAAGAGGGGTATGCCCATGCCAATGACATTAAGCAGGACGCGTGAATG 26901
Oy      2909 lUaAspSerSerGlySerGlu 2915
Db      26902 AGGACTGTCAGGCTCCGAG 26921
Oy      26902 ----- 26902

```

RESULT 11  
 BX284647 227194 bp DNA linear HTG 04-MAR-2003  
 LOCUS BX284647

DEFINITION Homo sapiens chromosome 1 clone RP11-6L21, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION BX284647  
 VERSION BX284647.1 GI:28865245  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEPIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 227194)  
 Hall, R.  
 Direct Submission  
 Submitted (03-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 Center: Wellcome Trust Sanger Institute  
 Genome Center  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 Project Information  
 Center project name: b66L21  
 ----- Summary Statistics -----  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator ABI; 0% of reads  
 Chemistry: Dye-terminator ABI; 96% of reads  
 Chemistry: Dye-terminator Big Dye; 3% of reads  
 Consensus quality: 164598 bases at least Q40  
 Consensus quality: 164660 bases at least Q30  
 Consensus quality: 164693 bases at least Q20  
 Insert size: 227194; sum-of-contigs  
 Insert size: 168516; 3.2% error; agarose-fp  
 Quality coverage: 7.72x in Q20 bases; sum-of-contigs Quality coverage: 10.52x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 227194: contig of 227194 bp in length.  
 Location/Qualifiers  
 1. 227194  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone\_id="RP11-6L21"  
 /clone\_lib="RPC1-11.1"  
 1. 227194  
 /note="assembly\_fragment:02685  
 clone\_end:SP6  
 vector\_side:right  
 clone\_end:T7  
 vector\_side:left"  
 misc\_feature  
 1. 227194  
 /note="assembly\_fragment:02685  
 clone\_end:SP6  
 vector\_side:right  
 clone\_end:T7  
 vector\_side:left"  
 BASE COUNT 64124 a 52119 c 51230 g 59721 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 227194  
 Score: 12339.00 Matches: 2207  
 Percent Similarity: 36.92% Conservatives: 0  
 Best Local Similarity: 36.92% Mismatches: 8  
 Query Match: 79.38% Indels: 4967  
 DB: 2 Gaps: 32  
 US-09-916-849A-3 (1-2923) x BX284647 (1-227194)

QY 1 MetarSerpProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 DB 3330 ATGGGAAACCCGGGACACCGGGGTCCCTCCCAAGCCGCGCGCTGTGTGCTG 3389  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 DB 3390 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449  
 QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyTyrLeuCysProSer 60  
 DB 3450 GGCTCCAGGAGGAGGAGGCTTCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 3509  
 QY 61 SerAlaSerAlaLeuTyrLeuTyrThrSerArgCysArgAlaGlyTyrGlnLeuThr 80  
 DB 3510 TCAAGGTGAAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3569  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100  
 DB 3570 GGCCACCTGGTAACCCACAGATGGCTGAGGAGGCTTGTGCTGCTGCTGCTGCTG 3629  
 QY 101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyTyrIleGly 120  
 DB 3630 ATTCCCTTACCAACGCTCTGAGAGCTGCTGAGCTGCTGAGCTGCTGCTGCTGCTG 3689  
 QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnHisProCysLeuValAla 140  
 DB 3690 GGCCACCTTTCCTCCCAAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3749  
 QY 141 ProAlaGlnLeuArgCysGlnSerCysGlyLeuValAlaGlnAlaProGlyValArgAlaGly 160  
 DB 3750 CCACGCTCAATGCAATGCAATGCTGCAAGCTGCAAGGCTGCAAGGCTGCAAGGCTG 3809  
 QY 161 ArgSerProGlnGlnSerLeuGlyValArgArgValArgAlaValAsnThrAlaProGln 180  
 DB 3810 AGGTACCAAGAAAGTCCCTGGTGGCTGAGAAAGAAATGTAATTAACCCCTCCAG 3869  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnGlnProAlaGlyTyrProVal 200  
 DB 3870 TTCCAGCCCCCAAGTACCAAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3929  
 QY 201 AlaSerLeuAlaGlnAlaAspProAspGlnGlyGlnAlaGlyValArgLeuGlnTyrThrMet 220  
 DB 3930 GATCCCTGAGGAGGCTGATGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 3989  
 QY 221 AspAlaLeuPheAspSerArgSerAlaGlnPhePheSerLeuAspProValThrGlyVala 240  
 DB 3990 GATGCCCTCTTGTATAGCCGCTCCACCAAGTCTTCTCCGAGCCAGTCACTGCTGCA 4049  
 QY 241 ValThrThrAlaGlnGlnLeuAspArgGlnThrThrThrThrHisValPheArgValThr 260  
 DB 4050 GTAAACCAAGCCGAGAGGAGTGAATGTAAGACCAAGACCAAGCTTCAAGGCTCAG 4109  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrThrThrThrThrThrThr 280  
 DB 4110 GGGCAGGACCAAGGATGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCT 4169  
 QY 281 AspThrAspAspHisAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 DB 4170 GACACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCA 4229  
 QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro 320  
 DB 4230 AACCTGAGGAGTGGCTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4289  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyValProSerGlnValPhe 340  
 DB 4290 AATGCCAATATTCCTGACCGCTGCTGAGAGGCTGCTGAGAGGCTGCTGAGAGGCT 4349  
 QY 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnVal 360  
 DB 4350 GAGATGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4409  
 QY 361 GlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGlnArgAspProGlyProArgSer 380



```

Db      4410 GAATCTTACGATGACGGGTAGAGGCAAGTGACAGAGTCGGGACCCGGGCTCTCGAGAT 4469
QY      381  ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPheSer 400
Db      4470 ACCACAGCGCGGTCTTCTCTTCTGTGAGAGATGACAAATGATATCCCTCCAGTTTAACT 4529
QY      401  GluLysArgTyrValValGlnValArgGluAspValThrProGluValaProValLeuArg 420
Db      4530 GGAAGCGCATGTGTCCAGGTGAGGAGAGATGTACTCCAGGGGCCCAAGTACTCCGA 4589
QY      421  ValThrAlaSerAspArgAspLysSerAspAlaValValHisTyrSerIleMetSer 440
Db      4590 GTACACGCTCGGATGAGACCAAGGGAGCAATGCGGTGTGACATATGACATCAAGAT 4649
QY      441  GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGluValaLeuAspValValSer 460
Db      4650 GGCAATGCTCGGGGACAGATTTATCTGAGATGCCCAAGCTCGAGCTTGAGTGTGAGC 4709
QY      461  ProLeuAspTyrGluThrThrTyrGluTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db      4710 CCTCTTGACTATGAGACCAAGAGATCACCCCTACGGGTGAGCACAGATGGTGGC 4769
QY      481  ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db      4770 CGTCCCACTCTATATGTCTCTGCTGTGAGACAGTCAAGTCTCGATATCAACGAC 4829
QY      501  AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      4830 AATGCCCATCTCTCTCAGACACCCCTTCCAGGCTACTCTCTGAGAGTGTCCCTTA 4889
QY      521  GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db      4890 GGCTACTGGTCTCATATGTCTCAGGCTATCGACGCTGATGTGTGTGACAAATCCCGCTG 4949
QY      541  GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly 560
Db      4950 GAATACCGCTGCTGCGGGTGGAGCATGACTTCCCTTCAACATCAACATGGCACAGGC 5009
QY      561  TrpIleSerValAlaAlaGluLeuAspArgGluGluValaAspPheTyrSerPheGlyVal 580
Db      5010 TGGATCTCTGTGGCTGCTGAATCGACCGGAGGAAGTTGATTTCTACAGCTTTGGGGTA 5069
QY      581  GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      5070 GAAGCTCGAGACCATGGCACTCCAGCACTCACTGCTCGGCGAGTGTGAGCTGACTGTC 5129
QY      601  LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      5130 CTGAGTGTCAAGCAACATCAACCTTTACCAACAGAGTACACAGTGGGCTCAAT 5189
QY      621  GluAspAlaAlaValaGlyThrSerValValThrValSerAlaValaAspArgAspAlaHis 640
Db      5190 GAGAGTGCAGCTGTGGGACCAAGCGTGTGACGCTGTGACCGTGTACTCAT 5249
QY      641  SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      5250 AGTGTCACTACCTACAGATCAACAGTGGCAATATCTCGAAACCGCTTCTCATCAACAGC 5309
QY      661  GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArgGln 680
Db      5310 CAAGATGGTGGGTGGTGTATCCCTGTCCCTGCCATCGAGATCAAACTTGAGGGGAG 5369
QY      681  TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      5370 TATGTGTGGCTTACCGCTCCGATGGCACTCGGAGAGACCGGACAGATTTGGTG 5429
QY      701  AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      5430 AATGTACCGGACGCGCAACCACTCATGTCTGTCTTCAAGAGCTCCACATATACAGTGAAT 5489
QY      721  ValaGlnGluAspArgProAlaGlyThrThrValValaLeuIleSerAlaThrAspGluAsp 740

```

---

```

Db      5490 GTTAATGAGACCCGCGCGGACGACCAACGAGTGTGTGATGACGCGCACGAGTAGAGAC 5549
QY      741  ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      5550 ACGAGTGAATGCGCGCATCACTTCAATGAGAGACAGATCCCGCATGTCGCAATC 5609
QY      761  AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      5610 GATGCAGACAGGGGGCTGTCAACCAACAGGTGAGCTGAGCTTAAGAGACCAAGTCTT 5669
QY      781  TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrTyr 800
Db      5670 TACACCTGGCCATTAACGTCTCGGACATGGCATTTCCCAAGAAATCCGACACCACTAC 5729
QY      801  LeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db      5730 CTGAGATCTGTGTAGAGAGTGAATACAAATGCCCTCAAGTCTCTGAGAACCTCTAC 5789
QY      821  GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      5790 CAGGCGAGTGTATGAGAGATGTGCACCTTCACTAGCGTCTGCAATCTCAACCACT 5849
QY      841  AspArgAspSerGlyLeuAsnGlyArgValaPheTyrThrPheGlnGlyAspAspGly 860
Db      5850 GATCGTATCTTGACCTTAATGACGAGGCTCTTCAACCTTCAAGAGCGCAGATGGA 5909
QY      861  AspGluAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLeuAsp 880
Db      5910 GACGTGACTTATATGTATGATCAAGTCAGCAATCGTGCAACCTACGGAGGCTGAGT 5969
QY      881  ArgGluAsnValaIaGlnTyrValLeuArgAlaTyrAlaValaAspLysGlyMetProPro 900
Db      5970 CGAGAGACGTGGCCCAAGTATGTCTTGGGCAATATGACAGTGCACAGGGATGCCCA 6029
QY      901  AlaArgThrProMetGluValThrValThrValLeuAspValaAsnAspAsnProVal 920
Db      6030 GCCGCGACACCTAATGAGAGTCAAGTCACTGTGTGATGTGAATGACAAATCCCTCTGC 6089
QY      921  PheGluGlnAspGluPheAspValaPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      6090 TTGAGAGAGATGAGTTGATGTGTGTGTGAGAGAGAACACCCCATTTGGGCTAGCCGTG 6149
QY      941  AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      6150 GCCCGGTCACAGCACCTGACCCCGATGAGGACCAATGCCAGATATATGACCAAGT 6209
QY      961  ValGluGluAsnIleProGluValaPheGluLeuAspIlePheSerGlyGluLeuThrAla 980
Db      6210 GTGAGGGCAACATCCCTGAGGCTTTTCACTGAGCAATCTTCTCCGGGAGCTGACAGCC 6269
QY      981  LeuValaAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db      6270 CTGTGATGACTTGAAGTACAGAGACCGGCTGAGTACGTCCTGAGTCAAGGCACTGCA 6329
QY      1001  AlaProLeuValSerArgAlaThrValHisValaArgLeuLeuAspArgAsnAspAsnPro 1020
Db      6330 GCTCCTGTGAGACCGGGCTACAGTCCAGTCCGCTTGTGACCGCAATGACCAACCA 6389
QY      1021  ProValLeuGluAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db      6390 CCAAGTGTGGCAACTTGAATCTTTTCAACAACATATGTCAACATCCGCTCAAGAGC 6449
QY      1041  PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      6450 TTCCCTGGGGGTGCATTGGCCAGATACCTGCCATACCTGATATCTAGATAGTCTG 6509
QY      1061  ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      6510 ACTTACAGCTTTGAGCGGGGAATGAACTACAGCTGTCTCTGCTCAATGCTCCACAGGT 6569
QY      1081  GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db      6570 GAGCTGAAGCTAAGCCGCGCATGAGCAACACCGGCTCTGAGAGGCCATATGAGCGTG 6629

```

QY 1101 LeuValSer----- 1103  
 Db 6630 CTGGTGTTC-AGGTAAAGAAAGGCCAGATGCGCTGGGGTGGGGGTAGTCTCGCGGGGATG 6688  
 QY 1103 ----- 1103  
 Db 6689 GTCTGGGACACCACTGAGAGTAAAGGTGCGATCCAGAAAGCACTACAGATCCACTCCCT 6748  
 QY 1103 ----- 1103  
 Db 6749 GCCCAGTGCCTGGCAGAGACAGAGAGGGGCAAGGCCAGCTTGGGAAGAACCCCAAGAA 6808  
 QY 1103 ----- 1103  
 Db 6809 CCCTGGCGGCTGGTGAAGAGGCCCACTTCTGTGGCTGGGCTGGCTGGCTGGCTGGCTGGC 6868  
 QY 1103 ----- 1103  
 Db 6869 TGGCCCCCACTGTCTACCAAGCCTGTGAAGTGTGCGGAAAGTGTGTGAAGGCTGCCCCGAC 6928  
 QY 1103 ----- 1103  
 Db 6929 TATGCTGCCAAGATTGTGAAAAAGGGCGTGAAGCCTGGGGGGCCAGAGAGCTGG 6988  
 QY 1103 ----- 1103  
 Db 6989 GGGAAAGGAGAGAGAGACTGTGGCGAGAGAGAGAGGCGAGGCCCAAGGCCCTTCAACCCC 7048  
 QY 1103 ----- 1103  
 Db 7049 CATCAAGATATTTCTGCAAGTCTCAACAAGCCTTTGGCAAGACTGGGGACCAAGGGAGCAC 7108  
 QY 1103 ----- 1103  
 Db 7109 AAGGCTGCGGCAACCCAGGGGTGAGAGCCCTGTAATCTCTTTGGCCCCCTTTGGCTCC 7168  
 QY 1103 ----- 1103  
 Db 7169 TGGCTGAGCTGCAAGATCGGCTGACAGGACCTTGTCTAGCGCATCTTTTCCCTCCAGG 7228  
 QY 1103 ----- 1103  
 Db 7229 CAGAAAGAGCTGGCTGGCTCCCCCTTCAGAGCTCCACAGGAATGAGGCCAGAGCTC 7288  
 QY 1103 ----- 1103  
 Db 7289 ATTGTCTGTCAAAACAGACCCCACTTTCAGAGTCCGTTGGGTCTGGCTTGAAGAG 7348  
 QY 1103 ----- 1103  
 Db 7349 GGGCTGCAATGAAGCTGGGGGGGAGGCTGAGGCTGTGGGCCCCCTTGGCACTCATG 7408  
 QY 1103 ----- 1103  
 Db 7409 GGGCTGTTTAAAGGTGCCAGGCCCTTCTTCTAGTGTGACCAAGCCCAAGCTCCAG 7468  
 QY 1103 ----- 1103  
 Db 7469 GAATGAGTGTAGGACACAGACAGGCAAGTGAAGTGTGAGCTTGAAGGACAGAGACC 7528  
 QY 1103 ----- 1103  
 Db 7529 CAAGGACCTTATCTAGAGAGAGACAGAGTCCAGGGCGTGGAGCTGGCCGGAACCC 7588  
 QY 1103 ----- 1103  
 Db 7589 AAGCTGGGGGCAATCTGTGCTTGGCCAAAGCACTAGCCTTACAGACTCCTTGAACAGCAG 7648  
 QY 1103 ----- 1103  
 Db 7649 GGCTGGGGAGAGAGTGTGAGGGGGTGTGACAGTATTTCTGTCTTTTATTCAGCTG 7708

QY 1103 ----- 1103  
 Db 7709 TGGGAGCCAGAGGTAGACAGAGGAGAAACACACCAAGTTGTGATGCTTCCCC 7768  
 QY 1103 ----- 1103  
 Db 7769 AAGCTGGAAACACTGAGACCTCATAGAGGAGATGTGTGAGCTGACACCCAGGGGTGG 7828  
 QY 1103 ----- 1103  
 Db 7829 ATCAGTCAATCCCTCTTAAAGGAGCTTCAATTCATCAGGGAGGAGGAAAGAGAA 7888  
 QY 1103 ----- 1103  
 Db 7889 GGCTGGAGGCCCTTATCCACCCAGAGAGTGTGCTGAAGGGAGAGAGGGATAG 7948  
 QY 1103 ----- 1103  
 Db 7949 CCAGAGTCAGGGCGTGGCTCACTCAGGGTCTTGTGAGACAGATGCTCCCTGCCAC 8008  
 QY 1103 ----- 1103  
 Db 8009 CCTCAAGTAGCAACAGGAAACAGAAAGCAGAGCTTTTCAAGGACTGTGGCTAGGTT 8068  
 QY 1103 ----- 1103  
 Db 8069 CTGAGCTTCTGAGACTTCAAGTTTCCCTCCAGAGGCGGAAGAGAGGCTGAGCTCC 8128  
 QY 1103 ----- 1103  
 Db 8129 TGTGTCTCAACTGTGCTGGGCACTTGAACAAGTGGAAAGAGGTGGAAAGAGGCC 8188  
 QY 1103 ----- 1103  
 Db 8189 TCTCCCTGTGCTCTCAACACAGCTGCTTTGTGAGGTGCGCAGGCACTGGCAGA 8248  
 QY 1103 ----- 1103  
 Db 8249 GAAAGGCCCAATTAAAGCTTGTGCTCCACAGGAGAGAGCCTTGAACCTTGTGCC 8308  
 QY 1103 ----- 1103  
 Db 8309 TAGTCTAAGAAAGGCTCGGGGCAATCAAGAGCGGCTCTCTGGCCCCGAGGCTGG 8368  
 QY 1103 ----- 1103  
 Db 8369 CCCCACTTCCCTTTCGAGCCTTGAACCTTGAAGCACTCAACAGCAACCCAAAGTT 8428  
 QY 1103 ----- 1103  
 Db 8429 CAGGAGGCCGGGTGAGGGGAGGAGTGACTCCACTGCCCCCTTGTCTTCTCTGTGT 8488  
 QY 1103 ----- 1103  
 Db 8489 TCTTGTCTCTTCTTCTTCCCGGCCCTGACAGGAGCCCTGACAGCTGCTCTC 8548  
 QY 1103 ----- 1103  
 Db 8549 TCCGGTGCCTTCCCTGCTCTGCGGCTCTCCGCTCTCCACTTGGCCATTTGCTGGCT 8608  
 QY 1103 ----- 1103  
 Db 8609 CCTTCTGAGGCTTGTGCTTCCCTGTTTGTGCTCCGTTGTGAGCTTCCACTCTTTC 8668  
 QY 1103 ----- 1103  
 Db 8669 CCTATGGTCAGAGTGGCCAGGCTGGCCAGATGGCTCCCTGAGGCTGAATTCCTCTC 8728  
 QY 1103 ----- 1103  
 Db 8729 GGCTCTTCTTCTCTGCTCAGACAGCAGAGCTCTGTTCATAGTGTGGGCAAGCTCCC 8788  
 QY 1103 ----- 1103

D	b	8789	AGAACAGACAGACCCCTCTTGGCCTCCAGTCTCCGCTGCACATCTGCCAGGGGAGAG	8848
Q	y	1103	-----	1103
D	b	8849	GCTGGAGGCTGACCCGGACCTTGGCACTTTTCAGGTGTCCCTGGCTAAGCTTGGGAC	8908
Q	y	1103	-----	1103
D	b	8909	AGTGGCATTTCTGGGGTCAGAGGCCATCCAGCTGTTCTTTTCTGTCCAGGAAGCA	8968
Q	y	1103	-----	1103
D	b	8969	GCTCAGCAACGTACTCCCAACCCCTGTTCCTCTCTGCTTGTGCACTGAGGGGTGACC	9028
Q	y	1103	-----	1103
D	b	9029	TAGGTCMAAGCAGGGGAGGGTGATATAGAAAGAGAGACAGTGGGTTGACTTTGGGA	9088
Q	y	1103	-----	1103
D	b	9089	TGGGGACAACAGCAGAGAGGGGCAAGTGGCCACAGATGTGTGACCAATGGGAAGCAACA	9148
Q	y	1103	-----	1103
D	b	9149	GAGCATGTCTGGAGAGGGGCCCAAGCGGGCTTAAGACATGGCGGTGGGGACACAGG	9208
Q	y	1103	-----	1103
D	b	9209	GAGAAGTGTGTTGTTCCAGGGCTGAGGGGTGGCAAAATTCTCCCGCTCCATGGCCTTGG	9268
Q	y	1103	-----	1103
D	b	9269	TTGCCCTCTCTCTTTAATCAGTTATGGGCACTGTGTGTGAGTCTTGGCTGCTTT	9328
Q	y	1103	-----	1103
D	b	9329	CTGCCTGAGGGAATATCCCGTCAAGGAATCATTAGCAGATATACAGGTATACAAG	9388
Q	y	1103	-----	1103
D	b	9389	CTCTGTCCAGCTAGCATCTGTGTGGGGGAAGTCTGAACGGAGAGAAAGTAGACC	9448
Q	y	1103	-----	1103
D	b	9449	TGATGATATAATCTGGGCAAGCTTCTGTGAGCAGTGAGCTAGGCGAGTTAGAACT	9508
Q	y	1103	-----	1103
D	b	9509	CAGAGACTGTGGCTGTGTCTGTGGAGGAGAACAGCCTGTGCAGCAGCTTTGAGAC	9568
Q	y	1103	-----	1103
D	b	9569	GAGAGCAGAGGAACAGAAGTGGGTGATGACAGCAGAGCTTGTGTGAGAGCTAGTAG	9628
Q	y	1103	-----	1103
D	b	9629	AATTAACAAGTATGAGGCTGAGGGTAGTGTGGCACTGGCTGAGGACATGGCTGAGAT	9688
Q	y	1103	-----	1103
D	b	9689	GGGGGTCTGCTTTCAGCCTTCCCGCTTCAGCTGAGCAACGCTGTGACAGTCCACC	9748
Q	y	1103	-----	1103
D	b	9749	GCAAGTAGTACACACAGATGCAATTTGTCCATGTCCCTTGCTACTCCCACCC	9808
Q	y	1103	-----	1103
D	b	9809	ACCCAGAAAATATGGCTCTGCCCGTACTGTAGGGTAAAGATTGCTTTTATATC	9868
Q	y	1103	-----	1103

---

D	b	9869	TAGTGGGCTCCCGCTGATGTGTGTTGTGTTGTGTTGGGACAGAGTCTCACTGTG	9928
Q	y	1103	-----	1103
D	b	9929	TCAACCAAGCTGAGTACAAATGGCGCATCTCGCTCACTGCAACTCTGCTCTGGGT	9988
Q	y	1103	-----	1103
D	b	9989	TCAAGCAATTCCTCGCTCAGGCTCCGTAGTATCTGGGACCTCTGTAGTAGTAGATT	10048
Q	y	1103	-----	1103
D	b	10049	ACAAGCAACCAACACAGCCCACTAATTTTGTATTTTGTAGAGACGGGTTTCAC	10108
Q	y	1103	-----	1103
D	b	10109	CATGTGTGATGTGATCTGAACTCTGACCTCTGATCAACCGGCTCAGCTCCCA	10168
Q	y	1103	-----	1103
D	b	10169	AAGTCTGGACTACAGGGGTGAGCACTGGCGCCCGCCCTGTATGTTTAAATCC	10228
Q	y	1103	-----	1103
D	b	10229	GTTTATCTTTGTGAGACTGAATAGATGTGATCACTGAGTGTTCGAACACC	10288
Q	y	1103	-----	1103
D	b	10289	TACTATGTGAAGCATTTGTAATAATAAATATGAATATGCAAGTAGCTCATGCTGT	10348
Q	y	1103	-----	1103
D	b	10349	AATCCAGTCTCTAGAAGCTGAGTAGAGATCACTTGAGGCCAGAAATTGAGACCA	10408
Q	y	1103	-----	1103
D	b	10409	GCCTGACAAACAGCAAGAACCCCATCTCTACACAAACAAATTAGCCAGGCATAGT	10468
Q	y	1103	-----	1103
D	b	10469	TGTAGCACTGTGTGTCTAGCTATTGGGAGTTGAGGTGAGAGATCCCTTGAGCCA	10528
Q	y	1103	-----	1103
D	b	10529	GAAGTCAGAGCTACAGTAGATGATGTGCCATGTCTTCAAGTATAGGACAGAG	10588
Q	y	1103	-----	1103
D	b	10589	AGTCTGTCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	10648
Q	y	1103	-----	1103
D	b	10649	CTCTCTTCCAGGCCCAAGCTTCAATTTTATAGTAAATAAATAATTCATTCATGTGTA	10708
Q	y	1103	-----	1103
D	b	10709	TGTGATTCAGGTGTATGATATTTCTGGGCTTAGGGTGTGTGAGTAAGACACC	10768
Q	y	1103	-----	1103
D	b	10769	AGCAAGTGTATTTAGTCTTAGCTGTATATAGTAGTGTGTGCTGGGGCATAGCA	10828
Q	y	1103	-----	1103
D	b	10829	GGCATTTGGGTGTAATGTTTGAAGCAAGTGGGTGATGAGTAGTAGAGCCCTTAGACCA	10888
Q	y	1103	-----	1103
D	b	10889	GAGAGAGACTGATCAGTGGGGCTCTGGGGTCTCAGAGTGAAGACTCCAGGCCCTT	10948
Q	y	1103	-----	1103
D	b	10949	CTATCTGTGCTTCCGGTCCCAACCCGACTCACTGTCTTCAGAGATGGGGCAGTGG	11008

QY 1103 ----- 1103  
Db 11009 TGATGATGCTAGTGTAAAGTGGCTCAGCTGATTTCTGGGGTGGGGCCCCCAGGCT 11068  
QY 1103 ----- 1103  
Db 11069 ACCCAACAGACCCCAAGAGAGTGGGGGGAAGATGCTTTGTGTGTGATGTCATGGG 11128  
QY 1103 ----- 1103  
Db 11129 GATAGCAGGCTCTGGGGCCCAAGACGATGGGGCTCAGTGGAGGGGGACTAGTGGCC 11188  
QY 1103 ----- 1103  
Db 11189 CACTGACCCAGTTTACCCTCTCCCCCAACCCCGCAACTCCCTGCTGTGGGGGAG 11248  
QY 1103 ----- 1103  
Db 11249 TGGGGAACACAGCCATGCTTGCATTAATGCTGTGTAGAGTGTAGAGGCTGAGGGGGCA 11308  
QY 1103 ----- 1103  
Db 11309 GGGGATGATTTTGTGGGTCTAGGATGCTCTGGGCCCTGTAGAAAAGTATGAAAAA 11368  
QY 1103 ----- 1103  
Db 11369 CAACATAGGGTTCCTCCGACAGGACACAGACTCTGGGGGAGAGGACAGAGCGGTGC 11428  
QY 1103 ----- 1103  
Db 11429 TTGGGACGGGAAAGTGCACAAACATTAGAGGTGTGCGTGCATCAGCAGCAAGAGTT 11488  
QY 1103 ----- 1103  
Db 11489 TTGGGAAAGGATGCAAAGAGACTGCATAATCTCATCCCTGAGCCCTGAGAGATGTTG 11548  
QY 1103 ----- 1103  
Db 11549 GCCCAGACGTGACATACCTGGCTCAGCACCTTATGCGAGTTGAGAAAGAGACCTTTC 11608  
QY 1103 ----- 1103  
Db 11609 TGGTGGGTGTGACAGAAATTCAGAGCTGAATTCAGGCCCAAGGCTGGCTCTGACTGTG 11668  
QY 1104 ----- 1104  
Db 11669 TCCCTCTCCACAGACGCGCTACACAGCGTGAACGCCCAAGTGGCGCTGGGTGACATC 11728  
QY 1120 TleThraBpGluMetLeuThrHisSerIleThraLeuArgLeuGluUasPheSerProGlu 11788  
Db 11729 ATCACCAGTGAATGCTCAGCCACAGCATCAGCTGCGCTGAGAGCAATGTCACCCGAG 11788  
QY 1140 ArgPheLeuSerProLeuLeuGluPheIleGluAlaValAlaIleThrLeuAlaThr 1159  
Db 11789 CGCTTCCTGACACATCTGAGGCTCTTCAATCAGGCGGTGACCGCAGCGCTGGCAAG 11848  
QY 1160 ProProAapHisValValValPheAanValGlnArgAapThrAapIleProGlyGlyHis 1179  
Db 11849 CCACCCGAGCACCGTGGTGTCTTCAAGTACAGCGGAGCACCGACCCCGGGGGCCAC 11908  
QY 1180 TleLeuAanValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPhe 1199  
Db 11909 ATCTCAACCTGAGGCTGTGTGGTGGCCAGCCCGCAGGGCCCGGGGGCGGCGCTTC 11968  
QY 1200 LeuProSerGluAapLeuGlnArgLeuThrLeuAanArgSerLeuLeuThrAlaIle 1219  
Db 11969 CTGCGCTCTGAGGACCTGACGAGGCGCTATACCTTCAACCGCAGCTCTGACGCGCATC 12028  
QY 1220 SerAlaGlnArgValLeuProPheAapAapAmIleCysLeuArgGluProCysGluAan 1239  
Db 12029 TCGGCAACAGCGGTGCTGCTTCGACGACATCTGCTGGGGAGCGCTCTCGAGAAAC 12088

QY 1240 TyrMetArgCysValSerValLeuArgPheAapSerSerAlaProPheIleAaSer 1259  
Db 12089 TACATGCGCTGTGCTGTGCGTGTGCGCTTCCATCTCCCGGCGCTTATGCTCTCTCC 12148  
QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
Db 12149 TCCGTGCTCTTCCGGCCCATCACCCTCGTGAAGGGCTGGCGTGGCGCTGCGCGCGG 12208  
QY 1280 PheThrGlyAapTyrCysGluThrGluValAapLeuCysTyrSerArgProCysGlyPro 1299  
Db 12209 TTCACGGGTGACTACTGCGAGACCGAGGTGAGACTCTCTGCTACTGCGGCGCTGTGGCC 12268  
QY 1300 HisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuArgAapGlyTyrThr 1319  
Db 12269 CAGGGCGCTGCGCAGCGCGAGGGCGGCTACCTGCTGTGTGATGGCTACAG 12328  
QY 1319 ----- 1319  
Db 12329 GGTGACCAAGGAGGAGGATCTCATGGGCCAGCCCTGGAAGCTGACTGTGTGCAAGC 12388  
QY 1319 ----- 1319  
Db 12389 ACAATCAGAACAAATGCTGGCGGCTCATTTCTTCCGAGTGAAGTGCACATCA 12448  
QY 1319 ----- 1319  
Db 12449 CTGAGAGTCATTAATCCGGCTGCTCCCTGACAGACCCCACTGAGGAGGAGAGTGC 12508  
QY 1319 ----- 1319  
Db 12509 GAGTGTCTTCTCCGTGTGTGACAGACACAGCAGTGCATTTGGGGCAGACTGG 12568  
QY 1319 ----- 1319  
Db 12569 CTCCAGACAGTGGCGCTGTGTGTGTATCAGCTTGGCGCTAGTACTGTGACCT 12628  
QY 1319 ----- 1319  
Db 12629 TGAACAAGTACTTAACCTTCTGCGCTCGGTTATTGAAATGAGATATAGACCTGAC 12688  
QY 1319 ----- 1319  
Db 12689 TTAGGTCAATAGTGTGTGTGAGATTAAATGATTAATTGATCTTGGAAACAGCTGG 12748  
QY 1319 ----- 1319  
Db 12749 TGGCATGAGAGGTGTGGCTTGATTTAAGGGAACGTGAGATTGGTGACTTGCTCTG 12808  
QY 1319 ----- 1319  
Db 12809 GTCTATCAACACCGGTTTAGAGGTGAGCTGGGCTTAGAACCTGTCTTCCCCCACTG 12868  
QY 1319 ----- 1319  
Db 12869 AGCGAGTTCCTTTCTGAGCTGACAGCAATCTTTCTCCAGGCAATTAAGGGAAGAC 12928  
QY 1319 ----- 1319  
Db 12929 ACAGAGAGCTGGGGGAATGAGAGGGGTGAGTTCTGAGATGGGGTCTCCAGAAACC 12988  
QY 1319 ----- 1319  
Db 12989 ATTGACCTGTATGGGCTGTGGCAGATAGCTCCGCTCCGCGCAGCTGTGTAAAGA 13048  
QY 1319 ----- 1319  
Db 13049 GGATCTGTGGTAGGCGGCTTCCCTGTGGCTGTGACCAAGTCTTTACATCTCTCT 13108  
QY 1319 ----- 1319  
Db 13109 CCTTGGATGCTACTGTCTGTCTTAATGCTCTGACCCCTTCCCTTCAAGGCTCTGTGT 13168  
QY 1319 ----- 1319

Db	13169	CTTCTCAGGAAACCCCCCTGCTCCCTTCCCTGTSGGCTTTCCTTCTAGATGCCAGGC	13228
OY	1319	-----	1319
Db	13229	CCTGCTCTTGTCTCTCCAGGGCCATGCCACTGGCTCTGCCAGAGCATAGGCAGGCC	13288
OY	1319	-----	1319
Db	13289	CATCTCTTCCCGCTGCATCTTCTGTCCAGAACTTGCAGACTTGACTTCTCTGTG	13348
OY	1319	-----	1319
Db	13349	CTGCTAGGGGTGGGGGGAAGGGAGGAGCGGGCTTGGGGGAGGGAGAAATCTTTGTC	13408
OY	1319	-----	1319
Db	13409	CAGAGAGATGTGGGGGGGTTGGAGGGGGGGGTTAGACAGCGAACAAACATTTCC	13468
OY	1319	-----	1319
Db	13469	ATGAAAACAGAGCCAGTCACTAGGCAAGAGACTGAAACTATCTTAGAATCTGGGA	13528
OY	1319	-----	1319
Db	13529	ATGGGGAGGGAGAGAAAGGTGATTTTCCATCTGATGTGGAGTTTCACTCTCTGT	13588
OY	1319	-----	1319
Db	13589	TGAATCTGTGTGTCTCTCAGGGCTCAGTTTCCACTTCAGCATGCCAGGTGCAG	13648
OY	1319	-----	1319
Db	13649	AGTCTTAATCTGACCCGAGGGCTCAGGTGGAGTCTTCCCATCTCTGGGGGC	13708
OY	1319	-----	1319
Db	13709	TCCCCAGTTCCTCTCATATCTTTCCATTAACCAAGCAAACTTAGTGAAGCTGTGC	13768
OY	1319	-----	1319
Db	13769	CCATCCCCAGGGAGGAACTGATCTCAGTGCTGAGAGGACAAAGACACAGTGGAA	13828
OY	1319	-----	1319
Db	13829	GGAGGGAGAGGAGCTCCAGCCCTCCCGCTGCTCGAGGCCCATCTGTGTGCAC	13888
OY	1319	-----	1319
Db	13889	ATTTAACCTCTTGGGGGCCAGACACCCCGAGTGTCTTGTAATGTGACATCGAAGGC	13948
OY	1319	-----	1319
Db	13949	TCCGGAGTCCGGGCTTACAAACCAATGTGCCAGAGGCTGTGTTATGAAACCAAA	14008
OY	1319	-----	1319
Db	14009	TGGGTAGGGAAGTGTCTGGGGACTCCAGAACTTGCAGGGGAGCATGCTGGAGCTG	14068
OY	1319	-----	1319
Db	14129	TCAGAGAAGATGGCAACGAGAGAAATGGGGATGCCCTGGAGTTTGGGAGTATTACG	14188
OY	1319	-----	1319
Db	14189	AGGCTCAGGGAGGCCATGGAGCTATCACTGGTGTGGGGGGTCTCAGTTCCCTCCT	14248
OY	1320	-----	1325
		glycylhistidylval	

Db	14249	GTCTCCAGAGTACTTGGTACTCACTGGCTTTCTCTTGTCCAGGTGACACTGTAGGTTG	14306
QY	1326	Ser1IaaArgSerGlyArgCy <sup>5</sup> ThrProGlyVal1Cy <sup>5</sup> IysAsnGly1Gly1ThrCy <sup>5</sup> Val1Asn	1345
Db	14309	AGTCTCGCTCCAGGCGCTTGACACCCCGGGGTCTCGACAGAAATGGGGGCACTGTGTCAAC	14366
QY	1346	LeuLeuVal1Gly1Gly1PheIysCy <sup>5</sup> AspCy <sup>5</sup> ProSerGly1AspPheGly1IysPro1Cy <sup>5</sup> Cys	1365
Db	14309	CTGCTGGTGGCGGGTTTCAGAGTGGATGGTCCCATCTGGAACTTCAGAAAGCCCTACTCGC	14428
QY	1366	GlnVal1Thr1ThrArgSerPheProAlaHisSerPhe11eThrPheArgGly1LeuArgGln	1385
Db	14429	CAGGTGACCAAGCGACACTTCCCGGACACTCTTATATACACTTTGGCGGCTGGCCACAG	14488
QY	1386	ArgPheHisPhePheThrLeuAlaLeuSer-----	1394
Db	14489	CGTTTCCACTTCAACCCCTGGGCCCTTC- GTGAGTGGCTGGGCACTGGGGGTGGAGTGGG	14544
QY	1394	-----	1394
Db	14548	CTTGTTGGGCACTGAGAGTGCCTCTCTGTTCTGTGTGAAGTAAAGTCAAGGACTGA	14607
QY	1394	-----	1394
Db	14608	AGCGTGAAGACGCGCTGGCAGACAGCAGAGTCTCTTGAAGTTGCTGTGGTCTGCC	14667
QY	1394	-----	1394
Db	14668	AGCAGATCTCCCTCCCTCGCTGTGGTGAAGTGGGGGTGCTGCATCCAGTGGGTAC	14727
QY	1395	-----PheAlaThrIysGlnArgAspGly	1402
Db	14728	CCCATCTCCCTGCCCCCATCCCAACTCTGTTCAGGTTTGGCCAAAGAGGGCGACGGG	14787
QY	1403	LeuLeuLeu1y <sup>1</sup> AsnGly1ArgPheAsnGly1y <sup>1</sup> HisAspPheVal1AlaLeuGlnVal11e	1422
Db	14788	TTCGTGTTGTACATATGGCGCTTTCATATGAAGACATGACTTTGTGGCCTCGAGGTATC	14847
QY	1423	GlnGlnGlnVal1GlnLeuThrPheSerAla-----	1432
Db	14848	CAGAGGAGGCTCAAGCTCACTTCTCTGC- AGGTATCACAAGTGGCCCCCATCTTGCC	14908
QY	1432	-----	1432
Db	14907	CATCTTCAAAGGCCCCCAAGTCTTCCAGGCCCTGACCCCAAGCCACTATCTATACGC	14966
QY	1432	-----	1432
Db	14967	AAATCTGGGCCCAAGCCCCAGACCACTGGCAACCCCAACCTTGCAANTTCTAGCCCTGTCA	15026
QY	1433	-----GlyGlnSerThr1Thr1ValSerProPhe	1442
Db	15027	GGCATTCAGACTCACTGTGTCCTTTCCCAAGGGGAGTCAACCAACCAACGAGTGTCCCATTC	15086
QY	1443	Val1ProGly1Gly1Val1SerAspGly1Gln1ThrPhe1SerThrVal1GlnLeuIys1y <sup>1</sup> ThrAsnIys	1462
Db	15087	GTGCCCGAGAGATCTAGTATGGCCAGTGGCACTACGGTGCACGTAAATACTAACATAAAG	15146
QY	1462	-----	1462
Db	15147	GTGGGTGTGAGAGGCAACAGAGGTTGGGGGTTCTGTTCTTATGGCTAAGGCTTACCCAGC	15206
QY	1462	-----	1462
Db	15207	CTTGAATGGCAATTGCCTCAAGGCTTGGAGGCTGTGTCAAGGCAATTCTGGCTAGAGAA	15266
QY	1462	-----	1462
Db	15267	AGGAGAGGGCTGGAGAGGTGCCACTTGTGTGGGCAATCCACCCCACTGGGCACTGTGC	15326
QY	1462	-----	1462
Db	15327	ATGCTCTCTGTTCACTAGGACGCGCAACTTGGCACTGAGGAAGGATGTATACATCA	15386

[illegible]

QY	1595	lajpheserlysgluCyseproleuaglYphaglYglYyserysalagln-----	1611
Db	16467	CGTTCAGCTCGAGGTGCCCTCGTGGGCTTTGGGGGCAAGAGCTCGCCAGGGTGAAGAGG	16522
QY	1611	-----	1611
Db	16527	GCGGCTGTAAAGAGCCACAGCCTGGGGGCATCAACAGTCTGGGAATCTGGCAAGGTTGGG	16588
QY	1611	-----	1611
Db	16587	GCAAGCACTGGGCAAGGCTCTCTGTAGCGGGGCTGTGGTGAAGAAAGCTGTCTGGGCAAG	16644
QY	1611	-----	1611
Db	16647	AGCCCTGAAGAGGCAACAGAGAGACAAAGGGGCACTAGAGAGCAGGCTTGGGCAACAG	16700
QY	1611	-----	1611
Db	16707	GTGGGGCGGTCCAGGGCAAGTATCGCACTTTGAGAGGGCGGGGCTGTAGAGGGAGTGGGC	16766
QY	1611	-----	1611
Db	16767	CTGCTCCCGGTGAGCGGCCACCGCTGAGCATCAAGCCCAAGGGCCTCTGAGGCTCC	16822
QY	1612	-----	1620
Db	16827	ACCGTCACAGTCTGCTTTCTCTGCTCCCGAGAAATGGCCCAATCAAGACATTCCTCG	16888
QY	1621	GlYseSerleuValAlaTTPHlsglYleuSerleuProIlleSerGlNProTPTlYleu	1640
Db	16887	GGCAGAGCGCTGGGTGGCTGGCAATGGCTCTCCCTGCCATTCCTCAACCGCTGGTACCTC	16944
QY	1641	SerleuMetPheArGTThArTArgGlnAlaAerGlyValleuLeuGlnAlaIleThrArGgly	1660
Db	16947	AGCTCATGTTCCGCAAGCGCCACAGGCCAGGTGTCTGTGTGAGAGCCATCAACAAGGGG	17000
QY	1661	ArgSerThrIleThrleu-----	1666
Db	17007	CGCAGCACATCAACCCCTACAGGTGATGATGAGAAAGGCGGCTGGCCTTGCCCATTA	17066
QY	1666	-----	1666
Db	17067	GGGCGCTGTAGCTTACGAGCGGCTGACAGAAAGTGGCTGGGCAAGTCTTGGGCAAGGGT	17122
QY	1666	-----	1666
Db	17127	GGGACATATAGAGCGCGCTGATCCGTTGGAAAGTCATGTCTGCCACTGTTGGGCTT	17188
QY	1666	-----	1666
Db	17187	GAGGAAATAAATAACGTTTCTCTCTAATGAGCTCTGAGCTGGGCAAGGATGGGTTTGAT	17244
QY	1666	-----	1666
Db	17247	GGTCAGAAAGGGGAACTGTAGACTGAGAAACAGTATTAAAGCCGCTGTACCTGT	17300
QY	1667	-----	1680
Db	17307	TTTCTTTCTCTCGTGTCTCAGCTTACGAGAGGGGCAACGTATGTCTGAGCTGTGAGGGGACA	17366
QY	1681	GlYleuGlnAlaSerSerleuAlyleuGlnProGlyAArgAlaMetAerGlyAAPTTPHle	1700
Db	17367	GGGCTTAAAGGCTTCTCTTCTCGTCTGAGGCCAGGGCCGGGCCAATGAAGGTGACTGGCAC	17422
QY	1701	HlslAglNleuAlaleuGlyAlaSerGlylyProGlyYHlslAlaleuSerPheAerP	1720
Db	17427	CATGCACAGCTGGCACTGGAGGCCAGGGGGGGGCCGATGCCATTCGTCTTCGAT	17488
QY	1721	TYrGlYglNglNArgAlaglYglYAlaMetleuGlyProArGleuHlslsglyleuHlslleuSer	1740
Db	17487	TATGGGCAAGAGAGGAGGGCAACTGGGCCCCGGCTGCATGTGTGCACCTGAGC	17544
QY	1741	AsnIleThrValglYglYIleProGlyProAlaaglYglYValAlaArGglyPheArGgly	1760





2004 ----- 2004  
Db 19766 CAATTTCTGCTAAGCCTCTGCGCGCACCAACCTGGTAATTTTTTTTTTTTTT 19825  
Qy 2004 ----- 2004  
Db 19826 TTTTTTTTGTATTTTAAAGAGACAGGGTTTACCATGTTGGCCAGTCTGGTCTCGAA 19885  
Qy 2004 ----- 2004  
Db 19886 CTCCTGACCTCATGATCCGCGCTGCGGCTCCCAAGTGTGGATTACAGGCGTGA 19945  
Qy 2004 ----- 2004  
Db 19946 CCACTGACCCCAACCAATTTTGAATTTTTTTTTTTTTTTTGAAGAGAGTCTGCTC 20005  
Qy 2004 ----- 2004  
Db 20006 TGTCAATCAGGCTAACATGCAATGAGCAATGATCTCGGCTCTGCGAACCTCCACCTGACAG 20065  
Qy 2004 ----- 2004  
Db 20066 TTCAAGCATTTCTCTACTGACCTTCCGAGTACGTGTACTACAGGCGTGTGCTACA 20125  
Qy 2004 ----- 2004  
Db 20126 TGCCAGCTAATTTTTTATATTTTTTAAAGAGAGAGGTTTACCATGTTGGCCAGGA 20185  
Qy 2004 ----- 2004  
Db 20186 TGCTCTCATCTCTTGACCTTGATGATCAGCCGCGCTCGGCTCCCAAGTGCAGATT 20245  
Qy 2004 ----- 2004  
Db 20246 CAGGGGTGAGCAACGCTGCCAGCCCAATCTTGACCTTTAACAAAAATGCTGAAAAAC 20305  
Qy 2004 ----- 2004  
Db 20306 CTGCACTGTCTTGAGATAGGCTGTGTCTAGAGATTCAAAAGATGATCAGAGGGTTCC 20365  
Qy 2004 ----- 2004  
Db 20366 AATCGTCACTCAACCAAGTCTGAGATTAGCCACCAAAAGGACAGATCTGGGG 20425  
Qy 2004 ----- 2004  
Db 20426 AGTAAAGGGGTGTGGCTGTGAACCCAGGACCCAGGATGGGGCTGGCGCTGGT 20485  
Qy 2004 ----- 2004  
Db 20486 CCTGGACTGGCAATGAGCAGGTTTATAGGACAGAGATGCGGAGACAAAGTGTGCG 20545  
Qy 2004 ----- 2004  
Db 20546 TGAAGACAGGACAGATGGGGGCTCAAGAGAGTGTATGCCCTGTTCCTTAAGGCCATG 20605  
Qy 2004 ----- 2004  
Db 20606 TGTGACATGATCTTGCGAGATTTTTTGGCCGTGGTGTCTACATACCTGTCTTCCGTTT 20665  
Qy 2004 ----- 2004  
Db 20666 CACCAAGGCGCTTGTGTGCTAGTACTGTCCCGGCCATGAGATGCTGGCTCCAGCAG 20725  
Qy 2004 ----- 2004  
Db 20726 AACCTCTCAGCCAGCTCAGGGGCTTCTGTGTGTCTCCCTAGAGCGCGCCCTTTGGCT 20785  
Qy 2005 ----- 2021  
Db 20786 TCCTTCCCCCAGGAGATGCTGTGCGCCACTGTGATGAGCAGAGGGGTGGCTCCCCCAA 20845

2021 snLeuPheAsnCyethrSerIleThrPheSerGluLeuLysGlyPhe----- 2036  
Db 20846 ACCCTTCACTGACCGTCCATCACTCTCAGAACTGAAGGCGTTCGTAAATGAACCCC 20905  
Qy 2036 ----- 2036  
Db 20906 CTCACTTCATCTTTTCCCTGTCTGCTGCTGAGTCTCACTTGCCCCCACTCCCATCTT 20965  
Qy 2036 ----- 2036  
Db 20966 TGAGAACGGGCTTCTGGAATTCAGCCTGTGTCTTCTGGGCTCCACGCTGAGAGGCC 21025  
Qy 2036 ----- 2036  
Db 21026 GTCTCTACCTCTGAGGCTCTTGTGCTCAGAGTCTCGCTCACTCTGCTCTCG 21085  
Qy 2037 ----- 2051  
Db 21086 CTCGCTCTGTCTGAGGCTGAGCGGCTACAGCGAAATGATAGGCTTAGACTCAGGGCGC 21145  
Qy 2052 SerGlnLeuAlaLeuLeuLysAsnAlaThrGlnIleThrAlaGlyTyrPheGly 2071  
Db 21146 TCCAGAGCTAGCCCTGCTCTGCGCAGCGCACAGCAGCAGCAGCTGGCTTCCGCGC 21205  
Qy 2072 SerAspValIleValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlnSerThrGln 2091  
Db 21206 AGCAGCTCAAGGTGGCTTACCACTGCGCCAGCGGCTGCGGCCACAGAGACCCAG 21265  
Qy 2092 ArgGlyPheGlyLeuSerAlaIleThrGlnAspValHisPheThrGlu----- 2106  
Db 21266 CGGGCTTGGGCTGTCTGCCACACAGACGTGACTTCACTGAGGTGGGCTTGAGAGA 21325  
Qy 2106 ----- 2106  
Db 21326 TGCAAGGCTGCTGTGTAGATAGGGGTCAATGTAGTGAAGCTGTGTATGCAACTGG 21385  
Qy 2106 ----- 2106  
Db 21386 GGGCAGAGGGGCGCTCCATCCACTCAAGAGAGCTCCCTGTGCGAGAGGCTTCCA 21445  
Qy 2106 ----- 2106  
Db 21446 TGAACCTGTGACCCCTGGGCCAGCCCTTCCCAACCTTTCATATGTAATATGTG 21505  
Qy 2106 ----- 2106  
Db 21506 TGTGGGCGATGTCTGACCCAGACAGAGGCTGCTGTGTATCAATGGCTGGAGAGA 21565  
Qy 2106 ----- 2106  
Db 21566 AGCAGTATCTGAGCATGTGTGTGGGTCCAGGACAGGGCTGGAGAGCTTTATGTAGA 21625  
Qy 2106 ----- 2106  
Db 21626 GAATGAGAGAGGGCTTAAGGGGAGAGCATCCAGTGACTGATGTGGCAATGGCGTCCA 21685  
Qy 2106 ----- 2106  
Db 21686 GGAGCATTTCAAGGAGGCAACAGTGAACAGAGGAGGGGCTGTGGGAGAGAGGA 21745  
Qy 2106 ----- 2106  
Db 21746 CTGGCCGGGACAGGCGGGGCTTCATGTGGGAGATGTCAGGTGTGGTGTGACACA 21805  
Qy 2107 ----- 2118  
Db 21806 CCAACCGTGAACCTTGCCCAACCCCAAGATCTGCTGGGGTGGGAGCGCCCTCTGGAAC 21865  
Qy 2118 PAlaAsnLysArgHisTyrGlnLeuIleGlnGlnThrGlnGlyGlyThrAlaTyrPheLeu 2138  
Db 21866 AGCCCAACAGGCGCCTGGGAGCTGATTCAGACAGAGGGGTGGCACCGGCTGGCTGCT 21925  
Qy 2138 uGlnHisTyrGlnLysAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSe 2158

```

Db 21926 CAGGACATAGAGGCTTACGCGCAGTGCCTTGCGCCAGAACATGCGGCACACTTACTAG 21985
Qy 2158 rProphetHrIIleValThrProAmIle----- 2167
Db 21986 CCCCTTACCAATGTCACGCCCAACATTTGTAAAGCTGTGCTGGGTTGGGAGGGGTT 22045
Qy 2167 ----- 2167
Db 22046 TGTGAGGAGATGCCGACAGAGCGGCTGTGCTGGGATCTGCTGCTCAAGGTCCG 22105
Qy 2168 -----ValIleSerValValArgLeuAspLysGlyAsnPheA 2180
Db 22106 ATCTGTACCATCCCTTCTTAGTCATCTCGGTAGTGGCTTGACAAAGGAACTTTG 22165
Qy 2180 IAGLYAlaLysLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAspLeuGlnT 2200
Db 22166 CTGGGGCCAAAGCTGGCCCGCTACAGAGCCCTGCGTGGGAGACGCCCGGACCTTGAGA 22225
Qy 2200 hTrHValIleLeuProGlnSerValPheArg----- 2210
Db 22226 CAACAGTCATTCCTGCTGAGTGTCTTTCAG-AGGTCAAGTGGTGGCATGATGAGTTG 22284
Qy 2210 ----- 2210
Db 22285 GGAAGTCGACCCCAAGTGTCTGTGCAGACTCCAAGAGACAGGCGCCAGCTAAAGTGCAC 22344
Qy 2211 -----GluThrProProValValArgProAlaGlyProGlyGluAlaGlnG 2226
Db 22345 AGTGTCCCTCCCAAGAGAGCGCCCGGTGTGTACAGGCCCGGAGGCCCGGAGGCCAAG 22404
Qy 2226 IuProGlnGluLeuAlaArgArgGlnArgArgIleProGluLeuSerGlnGlyGluAlaV 2246
Db 22405 AGCCGAGAGAGCTGGCAGCGGACAGCGACGCGCACCCGAGCTGACCGAGGTGAGGCTG 22464
Qy 2246 aAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProIleAsnTyrAspProA 2266
Db 22465 TGGCGAGGTATCATCTACACGACCCCTGGCGGCTACTGCTCATTAATGACCTTG 22524
Qy 2266 sPlyAsrSerLeu----- 2270
Db 22525 ACAAGCGACGCTTGAAGGTACAGAGGCTAAGGGGACAGGTGTGGGTAGGGGATGGGTGGGC 22584
Qy 2270 ----- 2270
Db 22585 GGTGAGTCTGAGGATGAGAGGGGGTGGGGGCTGTCCGCCAGTCATGTGACGTGG 22644
Qy 2271 -----ArgValProLysArgProIleIleAsnThrPro 2281
Db 22645 TGACTGTGCACCTGACCTGCCCAATCAAGAGTCCCAACCGCCGATCATCAACACACC 22704
Qy 2282 ValValSerIleSerValHisAspAspGluGluLeuProArgAlaLeuAspLysPro 2301
Db 22705 GTGGTGAACATCAAGGTGCATATGATGAGAGCTTCTGCCCGGGCCCTTGACAAACCC 22764
Qy 2302 ValThrValGlnPheArgLeuLeuGlnThrGlnGluArgThrLysProIleCysValPhe 2321
Db 22765 GTCAAGGTGACAGTTCGCCCTGTGAGACAGAGGAGCGGACCAAGCCCATCTGTGTCTTC 22824
Qy 2322 TrpAsnHisSerIleLeu----- 2327
Db 22825 TGAACCAATCAATCTGTGAGCTGCACTGCGCCCTCGCCCTCAGGCTTCGGGCTGAAA 22884
Qy 2327 ----- 2327
Db 22885 GTCCAGGCCCCGATGCTCACTCCTTGTCTCCCTGACCCGCTTCTTCAACAGAG 22944
Qy 2328 ValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAspGlnSer 2347
Db 22945 GTCAAGTGCACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 23004
Qy 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValIleMetAspValSerArg 2367

```

---

```

Db 23005 CAGCTCAGCTGCAGTGCACACACATGACAGAGCTTCGCTGTGCTCATGAGACGTTTCTCG 23064
Qy 2368 ArgGlu----- 2369
Db 23065 CGGAGGTGTGGGCCCAAGGGGACGTGCAGAGCCGTGGTGGGACCCAGGGGACGGGG 23124
Qy 2370 -----As 2370
Db 23125 CTGGGTGCTCAGGTCTGCTCCCTTCTTAATTCCTGTGGCCCCCTGCCACCTACTGTGACAA 23184
Qy 2370 nGlyGluIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaI 2390
Db 23185 TGGGAGATCTGTGCATGAAAGACATGACATAGGTGGCTTAAAGTCACTTGGCTGCG 23244
Qy 2390 aLeuLeuLeuThrPhePheLeuLeuThrLeuLeuArgIleLeuLysSerAsnGlnHisG 2410
Db 23245 CCTTGTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 23304
Qy 2410 YIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyI 2430
Db 23305 CATCCGAGCTAACCTGACAGCTGCTGGGCTGGCTGAGTGTCTTCTCTGGGAT 23364
Qy 2430 eAsnGlnAlaAspLeuPro----- 2436
Db 23365 CAACCAAGCTGACCTCTGTAGATGCTCTTACTGCCCAGAACTGTCCACCTTCTC 23424
Qy 2436 ----- 2436
Db 23425 AGCGCGCTCCCAAGCCCCCACTGGCAACCTGTCTCTGTGCACATGAACTTAAATAGG 23484
Qy 2436 ----- 2436
Db 23485 TGCCATAGTACAGCACTGGGCCAGAGTTTCTCTTGTGTGCTCCCGGATGCCAG 23544
Qy 2436 ----- 2436
Db 23545 CACCTGCTTGCCGACAGCTTCCCTGAGAGAGTCCCAACCCAGGCCCTCTCCAT 23604
Qy 2436 ----- 2436
Db 23605 GCCTGACCCGAGAGAGCTGTGCTGGGCGGGCCCCGTCCGTGACCTGCGCTGGCC 23664
Qy 2437 -----PheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysT 2453
Db 23665 TGGGCTCCTCAGTTTGGCTGCACAGTGCATGTGCATCTGTGCACTTCTGTACTGTGCA 23724
Qy 2453 hrPheSerTrpAlaLeuLeuGluAlaLeuHisIleuTyrArgAlaLeuThrGluValArg 2473
Db 23725 CTTTCTCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 23784
Qy 2473 sPValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheI 2493
Db 23785 ATGTCAACACCGGCCCAATGCGCTTACTACTATGCTGTGGGCTGGGGGCTGTGCTTCA 23844
Qy 2493 IeThr----- 2494
Db 23845 TCACAGATTACTCCACCATTCACAGTCTTGGGGTCCACATCCGTGGATCCACTTTGT 23904
Qy 2494 ----- 2494
Db 23905 GCCATGTTCTTCACACCAATACAGAGCCCTGAGGCCCAATCCCATGCCCCAGGCGG 23964
Qy 2494 ----- 2494
Db 23965 CTTAATTCAGAGGTGTCCTGTGTTTAACCAAGCTGTGACGCGGCCACCCAGGCGTCA 24024
Qy 2494 ----- 2494
Db 24025 CTCCTTAATCTGGGAGATCGGTAGGGGCGCATGTGTGGCAGAACTTTTCCATGCTT 24084
Qy 2494 ----- 2494
Db 24085 CATGCTGGGCTGTGTAGGCCCACTGTGCGGACGCTTAATTCACAGGCCCTCATATCAACC 24144

```

QY 2495 -----GlyLeuAlaValGlyLeuAspProGlnG 2504  
 Db 24145 CTCACAGCTCCGGTCTGTCTCCATGCTCAGGGGCTA-GCCGTGGGCTCGAGACCCGAGG 24203  
 QY 2504 LTTTGTGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIleTrpSerPheA 2524  
 Db 24204 GCTACGGGAACCTTGACTTGTGCTGCTCCATATGACAGCTCATCTGAGATTGG 24263  
 QY 2524 LAGlyProValAlaPheAlaValSer----- 2532  
 Db 24264 CTGGCCGGAGGCTTGTGCGCTCTGGTAGTGTCTAGCAGGTGGGTGGGTGCACCTGT 24323  
 QY 2532 ----- 2532  
 Db 24324 GGCCCTCTTTGCTGTCTCTGTGCTGGGGCGGGTGTCTCAGAGACTCCCTTATAGA 24383  
 QY 2532 ----- 2532  
 Db 24384 ACAGCTGAGGCCACGGGGCCCTGTGGGCTGGGTGGAAGCTGTTGTCCCACTGACACC 24443  
 QY 2533 -----MetSerValPheLeuTyrIleLeuAlaAlaArgAlaS 2545  
 Db 24444 CCATGCCCTTGCCTCCCTGCTAAGTAGTCTTCTGTACATCTGGGGGCCCGGGCCT 24503  
 QY 2545 erCysAlaAlaGlnArgGlnGlyPheGlnLysGly----- 2557  
 Db 24504 CCGTGTCTGCCACGGCGGAGGGCTTTGAGAAAGAGTCTGTGTAGATATAGGTTGGG 24563  
 QY 2557 ----- 2557  
 Db 24564 GTGCCGTGGCCATGGGACGACACAGCATGGAGGTCTCATGGCCGAGCTCATGGCCTGT 24623  
 QY 2558 -----ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT 2575  
 Db 24624 CCTATCCCTCAGC-TGGGGCTTGAGGCCCTCTTCGGCGCTCTCTGCTGTAGAGCA 24682  
 QY 2575 hrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheA 2595  
 Db 24683 CGTGGCTGGCTGCACTGTCTCTGTCAACAGCAGACCCCTCTTCACTACTCTTTG 24742  
 QY 2595 LThrCysAsnCysIle----- 2600  
 Db 24743 CTACCTGCAATTCATTCACAGTACTGCGCCAGCCTGTGGAGAGGACCTGGGCT 24802  
 QY 2601 -----GlnGlyProPheIleP 2606  
 Db 24803 GTGGATGCCGATATATGACACAGCCGTGCTGCTCTTGTGCTGCAAGGCCCTTACT 24862  
 QY 2606 heLeuSerTyrValIleLeuSerIleGlnValArgIleValAlaLeuLysLeuAlaCysSerA 2626  
 Db 24863 TCCTCTCTTATGTGTGCTTATGACAGAGGTCCGGAAGACATCAAGCTTGCCTGACGCC 24922  
 QY 2626 rglYAspProSerProAspProAlaLeuThrLysSerThrLeuTrpSer----- 2642  
 Db 24923 GCAAGCCCAAGCCCTGAGCCCTGTCTGACCAAGTCCACCTTGACCTCGGTGAGGAGC 24982  
 QY 2642 ----- 2642  
 Db 24983 CAGGGGTCTCAGAGGGCGGTGAAGGAGGAGGAGGAGGCCCAAGTAGCCTTACT 25042  
 QY 2642 ----- 2642  
 Db 25043 TTGGCCCACTCCCTTTCTCTTCTTCATCCCTTCTGAAAGGTGAAAGGAGAGGTG 25102  
 QY 2642 ----- 2642  
 Db 25103 AAATGCTGTGTTGGCTGTGGAGAGAGATGTTGAAAGGGTGTAGATTTGACTTGA 25162  
 QY 2642 ----- 2642  
 Db 25163 GGAGATGGGGCTGAGACAGATATATCAGTATATAGTCTTGCACCTGACATGAAATCTCTCA 25222

QY 2642 ----- 2642  
 Db 25223 TGAATTTAGTAAATTTTCTGTCTCCCAAGGCTCTCAGAGAGACATTTCTGTCTCAT 25282  
 QY 2642 ----- 2642  
 Db 25283 CTTTATACAGACCATTATATATAGTACAGATCCCTGGGGGACATGACGGGTAGACA 25342  
 QY 2642 ----- 2642  
 Db 25343 GCTGTGGAGAGCTGTGACCACTGCAGGCTCAATATGGGACAGCGGAGAGAGAGGC 25402  
 QY 2642 ----- 2642  
 Db 25403 CCTCTTTAGACCTGAGCATGTGGAAGTGAACTTAGGGCAAGTTCCCTCCACCTC 25462  
 QY 2642 ----- 2642  
 Db 25463 CTTCTTCCAGCTGGGGCAGCCATCCCACTTACTGACTCTGTCTGTTCCCTG 25522  
 QY 2643 -----SerTrpAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyA 2661  
 Db 25523 CCTAGTCTTACACTGCCCCAGCCCTTACGAGATGGCGGCTGTACAGCCCTTAGAGG 25582  
 QY 2661 pSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIleP 2681  
 Db 25583 ACTGGCCGGCTCTCTGACAGACCAAGTCGCTGGGGCAAGGTCAAGCCCAAGTATCC 25642  
 QY 2681 rPheLeuLeu----- 2684  
 Db 25643 CTTCTTCTGAGTGAATCCCGAGATGGAGGTGAGAGAGGAGAGGGGCCACG 25702  
 QY 2684 ----- 2684  
 Db 25703 CATGTGACCCAGGCCAGCCAGCTGTGGAGTTGAGAGACACACTGTGGCTGACGT 25762  
 QY 2684 ----- 2684  
 Db 25763 GGGGGCAGCTTGTGATTGAAGCTGTAAAGACCAACAGCAGAAACAGATCCCAAGGG 25822  
 QY 2684 ----- 2684  
 Db 25823 AGAGAGAGACTGGGACCTTGGGCAAGGGGCCAGGCTGACCCCTTCAGCATGTCTCATC 25882  
 QY 2685 -----ArgGlnGlySerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlnLysProG 2703  
 Db 25883 TTCATAGGAGAGAGTCCGCACTGAACCTGGCAAGGGCCCCCTGGGGGATCCAG 25942  
 QY 2703 LysLeuPheLeuGlnGlyGlnAspGlnGlnHisAsp----- 2715  
 Db 25943 GAGGCTGTCTGTGAAGTCAAGCCAGCAGCATGG-TGAGACACAGACGCTTGACCA 26001  
 QY 2715 ----- 2715  
 Db 26002 CCCAGAGGGCAGTTGGCTGTCTTACTGAAGTGGGTGAGAGGTGGCTGGGCTGT 26061  
 QY 2716 -----ProAspThrAspSerAspSerAspLeuSerLeuGlnA 2728  
 Db 26062 GATCTCTCCCTGAGCTCTTAATCTCTGACAGGACTCCGACAGTGACTGTCTTTAAGAG 26121  
 QY 2728 sPAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnG 2748  
 Db 26122 AGACCAAGTGGCTCTTATGCTCTTACCACTATCAGACAGTGAAGAGAGAGAAAGAGG 26181  
 QY 2748 LngLngLngLngLngLngAlaPheProGlyGlnGlnGlyTrpAspSerLeuLeuGlyProG 2768  
 Db 26182 AGGAGGAAGAGAGGCCCTCTTCCCTGAGAGCAGGGCTGGGATAGCCTTGGGGCTTG 26241  
 QY 2768 LysLysLysLeuProLeuHisSerThrProLys----- 2779  
 Db 26242 GAGCAGAGAGACTGCCCTTGACAGTACTCCCAAGGGTGGCCAGCACTGGGGCTGTGGC 26301  
 QY 2779 ----- 2779

Db	26302	CTTGGGGCCAGTGGAGAGACAGTGGGGCTCTGGGGTTCTTGGAGAGACAGACTGGGTGTGT	26361
Qy	2780	-----	AspGlyG 2782
Db	26362	GGCTCTCTGGCTATCTGGCTCTGTGGGCTCATCTTCTCTTCCCAACAGAGGGGG	26421
Qy	2782	LYPFGIYProGIYLYVAAlaProtrIrpProGIYAAPheGIYTHrTHrAlaVagIuSerS	2802
Db	26422	GCCCAAGGGCTCGGCAAGGCCCCCTTGGCAGAGACTTGGGACCAAGCAAAAGAGAGTA	26481
Qy	2802	ergLYaenGIYAlaProGIuGIuArGIeuarGIuAenGIYAAPAlaIuSerArGIuG	2822
Db	26482	GTGGCAACGGGGCCCCCTGAGAGAGCGCTGGGGAGAAATGAGATGCTGTCTCGAGAGG	26541
Qy	2822	LYSerIeunGIYProIeunProGIYSerSerAlaGIuPohIuIlyG-----	2836
Db	26542	GGTCTTAGGGCCCCCTTCCAGGCTCTTCTGCCAGCTCAAAAGTGAAGTGGGCAACC	26601
Qy	2836	-----	2836
Db	26602	CCACTGCGCAAGCTCCCTTACTGACGAGCCTCATCTCAATCTTCTGTGGCCGACCC	26661
Qy	2837	-----GlyIleuIuYsIlySlyCYLeuProTHrIleSerG	2849
Db	26662	TCAACAGCCCCGCCGCCCAAGGCAATCTTAAAGAAAGTGTCTGCCACATCAGGG	26721
Qy	2849	LYuYSerSerIeunIeunArGIeunProIeunIuGIuCYeTHrGIYSerSerArGIYSerS	2869
Db	26722	AGAAAGACAGACTCTCTCGGCTCCCCCTTGAAGCAATGCAACAGGGGTCTTCCCGGGCTCT	26781
Qy	2869	erAlaSerGIYLYSerArGIYGIYProProArGIYProProArGIYInSerIeug	2889
Db	26782	CCGCTATGTAGAGGGAGGCGGGAGAGCCCCCTCCCGCCACCGCCCCGGCAGAGCTTC	26841
Qy	2889	IngIuGIuIeunAenGIYAlaIuIeProIleAlaIuIeSerIleYAlaGIYTHrValaIaspG	2909
Db	26842	AGGAGCAGCTGAAACGGGGTCAATGCCCATGCCATGAGCATCAAGCAGGACACGGTGAATG	26901
Qy	2909	LYuAPSerSerGIYSerGIu 2915	
Db	26902	AGCACTGTCAAGGCTCCAG 26921	
RESULT 12	AC106175/c		
LOCUS	AC106175	235320 bp	DNA linear HTG 13-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-7M10, WORKING DRAFT SEQUENCE, 4		
ACCESSION	AC106175		
VERSION	AC106175.5	GI:30578758	
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 235320)		
AUTHORS	Muzny, D., Marle, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Ayala, B., Becht, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhey, C., Burch, P., Burrell, K., Calderon, B., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., David, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinu, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Bayes, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finlay, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabies, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebreyes, E., Geer, K., Gilly, R., Grady, M., Guerra, W., Guevara, W.,		

TITLE	JOURNAL
AUTHORS	REFERENCE
TITLE	JOURNAL
Gharavate,P., Haaland,M., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Haylak,A., Henderson,N., Hernandez,J., Hollander,X., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huliyk,S., Huine,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivert,A., Kapathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowis,C., Kraft,C.L., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J., Meshewari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapa,P., Martin,K., Martin,R., Martinec,E., McWhinney,S., McLeod,M.P., McNelli,T.Z., Meenen,E., Mlgoavleyvic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgana,C., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervill,S., Neal,D., Newton,N., Nguyen,N., Norris,S., Nockelmeih,O., Okumuon,G., Olarpunagoon,A., Pal,S., Parks,K., Paternak,S., Paul,H., Perez,A., Perez,B., Plannoch,C., Plopper,F., Polndexter,A., Popovic,D., Primus,B., Pu,L.-L., Piazzi,M., Quiroz,J., Rachlin,E., Reeves,K., Regler,M.A., Reigh,R., Riley,B., Reilly,M., Ren,V., Renteria,M., Richards,S., Riggs,F., Rivera,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Saverly,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Stiller,C.D., Smajic,D., Speed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J., Steinhilber,M., Strong,R., Sutton,A., Svalek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valae,R., Vera,V., Villalona,D., Waldron,L., Walker,B., Wang,D., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleciyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,U., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Neiderhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G., and Gibbs,R.A.	Unpublished Direct Submission 2 (bases 1 to 235320) Worley,K.C. Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 235320) Rat Genome Sequencing Consortium. Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On May 13, 2003 this sequence version replaced g1:25007950. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/atlas/">http://www.hgsc.bcm.tmc.edu/projects/atlas/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a "contig-scaffold"). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: GAOA Center clone name: CH230-TN10 Assembly program: Atlas 3.0: Summary Statistics Consensus quality: 225681 bases at least Q40 Consensus quality: 227443 bases at least Q30 Consensus quality: 229154 bases at least Q20	

Estimated insert size: 234593; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 231058: contig of 231058 bp in length  
\* 231059 231158: gap of unknown length  
\* 231159 232689: contig of 1531 bp in length  
\* 232690 232789: gap of unknown length  
\* 232790 233881: contig of 1092 bp in length  
\* 233882 233982: gap of unknown length  
\* 233982 235320: contig of 1339 bp in length.

## FEATURES

source  
1. 235320  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-7N10"  
misc\_feature  
1. 1015  
/note="wgs\_contig"

BASE COUNT 59423 a 53960 c 55203 g 61642 t 5092 others  
ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 235320  
Score: 12001.00 Matches: 2789  
Percent Similarity: 38.64% Conservative: 47  
Best Local Similarity: 38.00% Mismatches: 75  
Query Match: 77.20% Indels: 4436  
DB: 2 Gaps: 33

US-09-916-849a-3 (1-2923) x AC106175 (1-235320)

QY 1 MetArgSerProAlaThrGlyValProLeuProThProProProLeuLeuLeu 20  
DB 222169 ATCCGAGCGCGGCTGCGCAGCGCCCTCCCAACCGCACTGCTGCGCTACTGCTG 222110  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
DB 222109 TTCTGCTGCTGCGCGCGCTGCGCACTAGAGATCAGTGAGGCGCTGCTCTCTG 222050  
QY 41 GlySerArgGlyArgGlySerSerGlyValAcylalProMetGlyTyrLeuCysProSer 60  
DB 222049 GGGTCGGGGGAGCGCGGCTCTCTGGGGCGCTGCGCCCGTGAGGCTGCTGCGCAGCG 221990  
QY 61 SerAlaSerLeuLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyTyrGluLeuThr 80  
DB 221989 TCGGCTTCGAACTCTGCGCTCTACCGACCGCTGAGAGATCTAGGCACTTAGCTGAC 221930  
QY 81 GlyHisLeuValProHisAspGlyLeuArgValTyrCysArgProGluSerGluAlaHis 100  
DB 221929 GGGCACTGCTGCGCCATTCAGATGCGCTTAGGGTTTGATGTCAGATACAGGGGCTCAT 221870  
QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
DB 221869 ATCCCTCTTCGCGCATCTCTGAGAGGCTGCGCTGAGGCTGCTGCTGCTGCTGCTG 221810  
QY 121 GlyHisLeuSerProGluGlyValLeuThrLeuProGluGluHisProCysLeuValAla 140  
DB 221809 GGAACCTTTCTCCACAGGCGAGGCTGACCTCTGCTCAAGAACACCTTGTCTTAAGGCC 221750  
QY 141 ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyIle 160  
DB 221749 CCACGGCTCAGATGCGCAGTCTTTCAGAGCTGCGCAGGCGCCACGAGGCTCAGGGCTGAG 221690

QY 161 ArgSerProGluGluSerLeuGlyValArgArgValAspAsnValAsnThrAlaProGln 180  
DB 221689 GGATCAGCAGAGGAGATCTATGGGTGGCGCAGAGAAAGATGTGATACGCTCCGAG 221630  
QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
DB 221629 TTCAGGCTCCCGAGCTACAGAGCGCACAGTCCCGAGAACAGCGCTGCTGCTCTGTT 221570  
QY 201 AlaSerLeuValAlaIleAspProAspGluGlyValAlaGlyArgLeuGluTyrThrMet 220  
DB 221569 GCTCTCTTAAGAGCATTAATGATCCGATAGAGGTGAGCGGGTCACTTAGACCAATAG 221510  
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
DB 221509 GATGCCCTCTTTGATAGCGCGCTCCCATCATTTCTTCTCTTGAGACCAATACATGCTGTT 221450  
QY 241 ValThrThrAlaGluGluLeuAspArgGluTyrIleSerThrHisValPheArgValThr 260  
DB 221449 GTCAACCAAGCTGAGAGACTGATCGGAGACCAAGACACCCATGCTTCAGGGTCAAC 221390  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 221389 GACACGATCATGATATGATCCCGACGAGTGCCTTGCCCATCTATCTTATGATGACT 221330  
QY 281 AspThrAsnAspHisAspProValPheGluGluGluGlyTyrGlySerLeuArgGlu 300  
DB 221329 GACACCAATGATCATGACCGCTGTTTATGACAGCAAGATATCAAGAGAGCTTCAGGGAG 221270  
QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 221269 AACCTGAGAGTGGCTATAGAGTGTCTTACAGTCAGAGGCGCACCGAGGAGATGCCCTCT 221210  
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGluValPhe 340  
DB 221209 AATGCAACATATCTATACCGCTGCTGAGAGGCGCTGAGGCGACCCCTCAAGAGTCTTT 221150  
QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
DB 221149 GAGATCGATCCGCTCGCTGGGGTATCCGAACCGCGCCCGTATGATCGGGAGAAAGTG 221090  
QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
DB 221089 GAATCTTCAAAATATATATGATGAGGAGAGTACCAAGGTCGGGACCCAGGCGCCAGAGAG 221030  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400  
DB 221029 TCACAGCTGCTGCTTCTCTCTGTGAGATGATATATGACAGCCGCCCATGTTAGC 220970  
QY 401 GluIleArgTyrValValGlnValArgGluAspValThrProGlyValaProValLeuArg 420  
DB 220969 GAGAGCGCTACGTGCTTCAAGTGGGAGAGATGACCCCAAGAGCCCGTCTCCGA 220910  
QY 421 ValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHisTyrSerTleMetSer 440  
DB 220909 GTCAAGGCTCGAGAGAGACAGAGGCGCAAGCGCTGAGTACATATATGATCAATGAGC 220850  
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
DB 220849 GCGCAACGCTCGGGAGACATCTTATCTGATCTGACATGAGACCTCGATGTGAGGC 220790  
QY 461 ProLeuAspTyrGluThrThrIleGlyGluTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
DB 220789 CCACCTCGACTATGAGACCAACAAAGATATACATCGCGAGATCCGGGCGAGATGTGGC 220730  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 220729 CGGCTTCACCTTCCAAAGGCTCCGGTCTAGTACCGTGCAAGTCTTACATCAATAGC 220670  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
DB 220669 AACGCGCCCATCTTTGTAGAGACCCCTTCCAGGCGCAGGCTCTTGAAGAGTACTTATA 220610  
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540



```

Db 220609 GCGTACCTGGTCTCTCACTGATCCAGCGATTGATGCTGATGCGGAGAGAACCCCGCTC 220550
QY 541 GtutYrAgtLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db 220549 GAATACACCTGCTGGCTGGGTTTGGGCAAGATCTCCCTTCACATTAACAAGGCGACAGGC 220490
QY 561 TTPJLeserValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db 220489 TGGATCTCGGTGGCGCGGAGTTGACCGGAAAGGTTGATTTCTACAGCTTCGGAAGTA 220430
QY 581 GtutAaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValIleThrVal 600
Db 220429 GAAGGCAAGGAGACAGGACCCCAACACCTCAGCGCTCAGCGAGGTCAAGTCAACATC 220370
QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGluProGluTyrThrValArgLeuAsn 620
Db 220369 CTGGATGTCATAGACAACAACCACTTTACGACCGCGAGTACACGGTGGGCTCAAC 220310
QY 621 GtutAspAlaAlaValGlyThrSerValValThrValIleSerAlaValAspArgAspAlaHis 640
Db 220309 GAGGACGGGCTGTGGGCAACAGTGTGTGACGGTGTGACCGTGTGATCGAGATCGTCAAC 220250
QY 641 SerValIleThrTyrGluGluIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db 220249 AGTGCATTAACCTACCAAGATCAACAGCGGCAACCCCGCAACCGCTCTCTCAACACAGC 220190
QY 661 GtutSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluValArgin 680
Db 220189 CAAGCGGGGGGTGGCTGTGTCTCCCTGGCTTACCGCTGAGTCAACAACCTCAAGCGGCAAG 220130
QY 681 TyrValIleuAlaValThrAlaSerAspGlyThrArgAsnThrAlaGluIleValVal 700
Db 220129 TATGTCGTGGCGGTGATCTGCTCGGATGGCAAGGCAAGACAAGCTCAAGATCGTGGTG 220070
QY 701 AsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 220069 AATGTCATGATGATGCAACCAACCATCGTCCGCTCCAGAGCTCCACATACAGGTGAT 220010
QY 721 ValAsnGluAspArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAsp 740
Db 220009 GTTATGAAAGACCGGCGACAGGACCAACAGTGTGTGATGATGCTCAACAGACAGAC 219950
QY 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGluPheArgIle 760
Db 219949 ACAGGGGAGAAATGCCGAATCACTTATGAGAGACAGATCCCTCAAGTTCGCAATC 219890
QY 761 AspAlaAspThrGlyAlaValIleThrThrGluAlaGluLeuAspTyrGluAspGluValSer 780
Db 219889 GATGACAGACATGGGGCGCTCCACCAACGCTGAGCTGAGCTAAGAAATCAAGGTGTCT 219830
QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGluIleSerAspThrThrTyr 800
Db 219829 TACACCTCGGCATACCGCTCGGACAAATGSCATCCCCCAAGATCCGACAAACCTAAC 219770
QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGluPheLeuArgAspSerTyr 820
Db 219769 CTGAGAGATCTCTGTAAATGACGTGAATGCAATGCCCCCAAGTCTCAAGAGATTCCTAAC 219710
QY 821 GtutGlySerValTyrGluAspValProProPheThrSerValLeuGluIleSerAlaThr 840
Db 219709 CAGGGAGAGTCTAAGATGAGATGTCACCTTCACACAGGCTGCTCAGATCTCAACCAAC 219650
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGluGlyGlyAspAspGly 860
Db 219649 GATCCGCACTCGGCTCTAAGACGCAAGGTTTCTCAACCTTCACAGAGAGGAGCAATGGA 219590
QY 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 219589 GATGTGTGATTTATGTAGAGTCCACATCAGGACATCGTGCACACTCGGAGGCTGGAT 219530
QY 881 ArgGluAsnValAlaGluThrValLeuArgAlaTyrAlaValAspGlyGlyMetProPro 900

```

---

```

Db 219529 CGTGAAGATGTGGCCCACTACATCTTGAGAGCGTACCGGTGGACAAGAGGATCCGCCA 219470
QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db 219469 GCGGCGACACCCATGAGAGATGACAGTACTGTACTGATGATGAAATGACAAATCCCGCTC 219410
QY 921 PheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 219409 TTTGAACAGATGATGATTGATGATTTGTGGAAAGAAACGCCCATCGACTGGCTGTG 219350
QY 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGluIleMetTyrGluIle 960
Db 219349 GCGCGGTGACAGCGCACCGCAACCAAGAGAGGACATGACAGATCACTTACAGATA 219290
QY 961 ValGluGluAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla 980
Db 219289 GTGAGGGCAATATCTCTGAGGTCTTTCAGCTGACATCTTCTGGGGCACTAACGGCC 219230
QY 981 LeuValAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGluAlaThrSer 1000
Db 219229 CTGTAAGATTTGACTACAGAGACCGAAGTGAATATATCTCGTCAATCCAGCTACGCT 219170
QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020
Db 219169 GCTCCCTGTGTGACAGGCTTACGTCTCATGTCTCGCTCTCGACCGCAATGATTAACCG 219110
QY 1021 ProValLeuGluValAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040
Db 219109 CCACTGCTGGGCAACTTGAATCTTTCAACAACATATGTCACCAACCGCTCAAGACAGC 219050
QY 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu 1060
Db 219049 TTCCCTGGGGGTGTATAGCGCGGTGCTGCTGCCATGACCCGATATCTCAGACAGCTG 218990
QY 1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerIleuValLeuLeuAsnAlaSerThrGly 1080
Db 218989 ACGTACAGCTTGTGCGAGAGAAATGAATCAAGCTGTGTCTCAATATGCTCCACCGGT 218930
QY 1081 GtutLeuValSerLeuSerAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db 218929 GAGCTGAGATGAGACCGGCGACTGGAACAACCGGCCCTGGAAGCATATAGTGTG 218870
QY 1101 LeuVal----- 1102
Db 218869 CTGGTGTCAAGTAAGAAATGCCAGTGAATGTTGGGTGGAAATAGTCTTTAGGAGAG 218810
QY 1102 ----- 1102
Db 218809 TCTTGAAGCCATGAGGTGTGTGGCTTCCAAAGAGGCTGTGTGACCCAGCTCTCCAC 218750
QY 1102 ----- 1102
Db 218749 CCTGTGATAGTCTCAAGATTTGTGAATCAGAAACCTGGCTGAGAAAGAAACCCAGAA 218690
QY 1102 ----- 1102
Db 218689 TCTGAGAGCTGSGCAGAGCCCGCTGTGCTGCCAGGCTGAGCTGGGCTGTTGTTG 218630
QY 1102 ----- 1102
Db 218629 GCTGCCCCCTGCTTACAGACTGTGACGTGTGTGGGAGAGAGGTTGTGAGCTGCTGGC 218570
QY 1102 ----- 1102
Db 218569 TCTGTGCTGCAAGAAATGTGAAAAAAGAGAGGCGGTGGAAGCTTGAAGGCCCA 218510
QY 1102 ----- 1102
Db 218509 GAGCTGGAGAGAGAGAGAGGAGCTGGAGAGAGAAAGGAGAGGAGGAGGAGGAG 218450
QY 1102 ----- 1102
Db 218449 AGACCTTGGCTGCGAGGAGATGGGGTGGAGAGGAGGCAATCGGTTCTCAGACCTTG 218390

```

QY	1102	-----	1102	QY	1102	-----	1102
Db	218389	TCACCTCATGGAACCTTGATCTCACACAGCTTTGGCAGGTGAGAGGCTACCAAG	218330	Db	217309	GCTTAGCCTTTGCTTCTCACACAGGACACAGAGCCCACTGACAACTCTAGATAG	217250
QY	1102	-----	1102	QY	1102	-----	1102
Db	218339	ACACAGATTCCACGAAGGGGGTCAGAGACCTGCGCTCTCTCCCACTGTGTTTC	218270	Db	217249	GGAAAGCCTGGCGCAGCTAGGGCCACTGAGCTGCTTTCTGCCCCCTGCTCC	217190
QY	1102	-----	1102	QY	1102	-----	1102
Db	218269	TCAGTAGCTGAGAACTGACTTGACTTGTCTAGACATTCCTTTCCCTCCAGGC	218210	Db	217189	CCATCAGCCTGACCTTAGAGTCATGACCAAGGCAACCCAAATTGAGAGGCTGGG	217130
QY	1102	-----	1102	QY	1102	-----	1102
Db	218209	AGAAAGCCTGCTTGCTGCCCTTCAGAGCTCCAGGCAAGAGTAGACAGCTTC	218150	Db	217129	TGAGAGGACTGAGGCCACATGCTTCTTGTCTCTCTGAGTTCTTGTCTCTCT	217070
QY	1102	-----	1102	QY	1102	-----	1102
Db	218149	ATTGCTCTGTCCAGACAGACCTCACTGCCCTGAGTTCTTGGTGGGGTCTGGCTTGGG	218090	Db	217069	TCCCTCCACCCCTGCCATGGGGACCCCTCTGCACTGCTGTCTGTCACTTCCCGAG	217010
QY	1102	-----	1102	QY	1102	-----	1102
Db	218089	GAAAGGCTGCACAAAGCCAGGCGAGATGACCTTTTGGGCTCCCCCAGCTGGCT	218030	Db	217009	CTTCTGTGCTTCTTCTTATAGCTTCTTATAGCTTCTCAGCTGCTGCTCTTGTCTTCCCTGTTTC	216950
QY	1102	-----	1102	QY	1102	-----	1102
Db	218029	TGTTCTTGGGGTGTAGATCTTCTCTGTGACATGATGGCATGGCTTCAGGAGAACAGA	217970	Db	216949	TCTATATTTAGATTCCTGTCTTCTCTCTCTCACTGTGGCTGAGCTGCCAGGCTGGG	216890
QY	1102	-----	1102	QY	1102	-----	1102
Db	217969	CACGGGGTTCAGAGTCACTAGGACTATCTCTTGTGAGGGCAGAGACTCCAGGGCCCTG	217910	Db	216889	CTCCATGCTCTGCAAGCCCAATGTCCGGCTGCTATCTCTTCAGACACACAGAGA	216830
QY	1102	-----	1102	QY	1102	-----	1102
Db	217909	ACTTGTGAGAGAGAACAGATTTCTAGGGGGGTGGAGCTGGCCAGAACCCAGCCTGG	217850	Db	216829	CCTGCCAAGGCTGCTCACTCTCCAGACAGACAGTCTTATCGAAGAGCAAGTTG	216770
QY	1102	-----	1102	QY	1102	-----	1102
Db	217849	GCATTCCTGCACTTGTCAAGGCTCAGCCCCCAGACAGAGAGATCGGGAGGGAAG	217790	Db	216769	GGAAGCTGACTGTGGCTCTTCCCTGTGACATGCTGAGGCTTAGACCAAGTATAT	216710
QY	1102	-----	1102	QY	1102	-----	1102
Db	217789	CAGACTGGGTGATATATTTATTCCTTCTGTCACTCTGAGATGAGAACAGAGCGG	217730	Db	216709	TGCTGGGTCAAGCCTACCCAGCTCTCTTCTTAAAAAAGTCAGGCCACTCTCTATTC	216650
QY	1102	-----	1102	QY	1102	-----	1102
Db	217729	GAGGGGGCAGACTCAGTAGCCAGACCTTTCCTGAGTCTGGGAGTTGGGGCTGGAGAGG	217670	Db	216649	CCAACTGCTTTCGGAAGAAAGACATGAGAGAGAGAGACATGAGTTGTTGCTG	216590
QY	1102	-----	1102	QY	1102	-----	1102
Db	217669	AACTGTATGTTTTTCAGAGGGGTGAGACTATGTGGCTTACTCCAGGCAATCTTACACATC	217610	Db	216589	GGATGGGACATGACAGAGGGGCAAGTGGCCACAGATGTGGCCAGATGGGGAGCAGC	216530
QY	1102	-----	1102	QY	1102	-----	1102
Db	217609	AGGAGGGGAGATGAGCTTCGCCCTTCTGCTCATACTTCTTACTGACAGACACTT	217550	Db	216529	AAAGCATGCGGGCAGAGGGCACAATCAGGCTTGGGCTGTGTAAACATGCGGAGTG	216470
QY	1102	-----	1102	QY	1102	-----	1102
Db	217549	CCGTATCCACCCCTGAGGTGGCTGCMAAGACCAAGGAGAAAGAGCTTCCAAAGCTATG	217490	Db	216469	TGGCTTCCAGGACCAAGAGATGCAAAAGATTTCTTGTCTTGTGTTTCAAGAGCACTGT	216410
QY	1102	-----	1102	QY	1102	-----	1102
Db	217489	GCTTGTCAAGCTGTAAACCACTGAAGCCTCAGTTCCCTCCAGAGAGAAAGTAAAGA	217430	Db	216409	TTGTAGTCTCCCTGTGTCTTCTGTCTGTCAGAGGGAAATCCCATTCAGAAAACACTGGGC	216350
QY	1102	-----	1102	QY	1102	-----	1102
Db	217429	GCAACCTCACTGAGCTGGCAGAGAGGCAAGAGCGGCTCTCTCCCTCCGCTTCCCA	217370	Db	216349	TGTAACTTCACTCTGAGCCCAAGTCAGGCTTCTGACAGAGGCAAAAGAGTGTGATC	216290
QY	1102	-----	1102	QY	1102	-----	1102
Db	217369	TCCACAGCCTTCAGTGTGATTCGGGCTAGATTCTTAGGGCCCGGTGTGAGAGCAGCTCT	217310	Db	216289	ACCATGTGACCATGAGGAGGCTGAGGGCTTAGAGGGCTGTGACTGTCTTGTGTA	216230
QY	1102	-----	1102	QY	1102	-----	1102

Db	216229	CCCTGGGCAAGCAGGAGGACGAGGTGAGCAGCACTATGGCAGCAGGCTTGTGAGAA	216170
Qy	1102	-----	1102
Db	216169	CAAGCAGATCACAACAGAGAGGTTGAAGGCAAGTAAAGCGCAAGCCAGGCAACAATG	216110
Qy	1102	-----	1102
Db	216109	AGATGGGTACTGCGCTTCAGTATGCCAAGCCTTGAAGTGTGACATATACAGAAATAGG	216050
Qy	1102	-----	1102
Db	216049	GTCCCATATCGCTTTTCTACTTCTGCGCTCAGACGACACTGCTCAAGTCCCATCCACTTA	215990
Qy	1102	-----	1102
Db	215989	GATGTCAATTTTGCTTCTGATTGAGAGTTTGCTCTCTCTCTCTCTCTCTCTCTCTCTC	215930
Qy	1102	-----	1102
Db	215929	TC	215870
Qy	1102	-----	1102
Db	215869	AGATGAATTTTTCTGTGTACTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	215810
Qy	1102	-----	1102
Db	215809	CCCTGAACCTCAGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	215750
Qy	1102	-----	1102
Db	215749	CACCAACCTAGTTCGCTGCCACGTTTTTAAGATGCTTTTATGTAAAGGATTCATC	215690
Qy	1102	-----	1102
Db	215689	AGTGTGTAGTTTTTAACTAGGAATACCCCTATACACCTTCTAAGCGTTTCCCTTTCT	215630
Qy	1102	-----	1102
Db	215629	TTGCTCCAGTTTTCAATTTTAACCTTAGTGCATTCGCTGTGTATGAACATTCAGGTGT	215570
Qy	1102	-----	1102
Db	215569	GCATGTACTTCTGAGCCTAGGTGCGTGGGTGAAGTTGGGCAACGTCGAATGTGTAGCT	215510
Qy	1102	-----	1102
Db	215509	GTTTGTAGTGAATGCGCTGCTAGAGGGTATGTGACGACAGACTTTGGGTACAATTGGT	215450
Qy	1102	-----	1102
Db	215449	GAGCCAGTGAATGAATTCCTTAAGAGGCCCTAGTGCAGAGAGAGAGACGCCAGT	215390
Qy	1102	-----	1102
Db	215389	GGGGCTCAGGGTGAAGAGCTGCCGCTCTCTTAATTTTACACAGATCCACCCCAACCC	215330
Qy	1102	-----	1102
Db	215329	CCCCACACCCTTTCTCAGACTTGGGGCACTGGGCGGTGTGCACATGGCCAGATGGCTC	215270
Qy	1102	-----	1102
Db	215269	AGCTGATTTCTTGGGCTGGGAGGCCAGGCTGCCAGAGTTGATGAGAGAGGCTCGG	215210
Qy	1102	-----	1102
Db	215209	TGTAAAGAGTGTGTGCATGGGGCTAGCAGGTTTGACCGAGACAGTATAGAGCTGAGTG	215150
Qy	1102	-----	1102

Db	215149	CAAGGGAAACGAGCTCAGTGGAAACATGTTTCCAGAAACGATTCGTGCACACCCAAAC	215090
QY	1102	-----	1102
Db	215089	CCGCCACAGCAGCGTGAAGAGCGGGATTATCGTGGGTTGAAGGTGAACGGCAGAGTTC	215030
QY	1102	-----	1102
Db	215029	TCTCTGGTCTCTGGGCCCTCTAGAAAAGAAAGAAAGAAAGCCGTTTGGACAGG	214970
QY	1102	-----	1102
Db	214969	GAAATCCAAGCAGTCGACGGGGACAGTGTATACGAATTAGGGTGAAGCAGAAACACT	214910
QY	1102	-----	1102
Db	214909	GCAGGCAGTAGTTACACGCTTCATGCAAGAGTGTGGGAAACACAAAGAACGTTTT	214850
QY	1102	-----	1102
Db	214849	CGCAATGTCCTTTTCTTAAGCACAATAGACGTAGCCCTCAAGAGAACGTGTCCA	214790
QY	1102	-----	1102
Db	214789	CAAGTGAACAGCTCATTTGGCCACTTATGCACTAGAAAAAGTTTTCTTGGGTACC	214730
QY	1102	-----	1102
Db	214729	TTGCCGATATAGGGCTAGATTTTAGCCCTGTGCACTTCTATTCGCCATGATCTCTTT	214670
QY	1102	-----	1102
Db	214669	ATGAGTCTCCCTTGTGTCCTCTCTCTGCGCCCTCCCTCTTATGACTCTTCTTC	214610
QY	1102	-----	1102
Db	214609	CTGGTCTCTCTCTCCCGGCCCTCCCTTGTATCTCCGTTTCTCTAGTCTCTTACT	214550
QY	1102	-----	1102
Db	214549	CGTATGCTCTCTCTCTGGTGTCTCCCTTTCAGATCTTCTCCGTGTGTCTCCC	214490
QY	1102	-----	1102
Db	214489	TCCTGACAGACCTCCCTCTCTCTCCGTTTCTTGACTCTCTGACAAACCTTAC	214430
QY	1102	-----	1102
Db	214429	TCCTCTCTGCTCTTAAAGCCAGGGCTTGGGCGCATCTGCTCTGACTTACCCCAAGGCTGA	214370
QY	1103	-----SerAspGlyValHisSerValThralagInCysAlaIle	1115
Db	214369	TGTTGATTGCGTGTGCTTCTCTTCAAGATGGTGTCAAGTGTGAAGCCAGTGTCACT	214310
QY	1115	ValGValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAs	1135
Db	214309	CCGGTCAACCATTCATCAACAGCGAGTCTCACACAGCATCAACGCTGCGGTGGAAGA	214250
QY	1135	pMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaI	1155
Db	214249	CATGTCTCCAGAAAGCTTTCGTCTCCCACTGCTGGAGCTCTTCATTCAGGCTGTGGCAC	214190
QY	1155	ATThrLeuAlaThrProProAspHisValValIlePheAsnValGlnArgAspThrAspAl	1175
Db	214189	CACATTTGGCCACACCCCAAGATCAACGAGTGTCTTCAATGTGACAGAGGATACGATGC	214130
QY	1175	aProGlyValHisIleLeuAsnValSerLeuSerValGlyValInProProGlyIProGlyI	1195
Db	214129	CCCAAGGGCCATATCTCAACGTAGAGCTGTCAAGTGGCCAGCCCCAGAGACCCGGGGGG	214070
QY	1195	YGlyIProProPheLeuProSerGluAspLeuGlnGlyValArgLeuTyrlLeuAsnArgSerIe	1215
Db	214069	TGGGCAACCCCTTTTACCTTCAGAGGATCTGCAAGAGGCGCTGTATCTCAACCGGAGGCTT	214010

QY 1215 uLeuThrAlaIleSerAlaGluArgValLeuProPheAspAspAlaIleCysLeuArgI 1235  
DB 214009 GCTTACCAGCACTCTCAAGCTCAAGGCGGTGCTCCCTTCGACGACACATTGGCTCGAGAA 213950  
QY 1235 uProCysGluAsnTYrMetArgCysValSerValLeuArgPheAspSerSerAlaProPh 1255  
DB 213949 ACCCTGCGAAGATTACATCGGTGTGTCCGTGCGCTTCGACTCCTCGCCGCTT 213890  
QY 1255 eIleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArg 1275  
DB 213889 TATCGCTTCCTTCGTGCTGCTTCGCGCCCATCCACCTCGGCGGTCTCGCTGTCG 213830  
QY 1275 GcysProProGlyPheThrGlyAspTYrCysGluThrGluValAlaPheLeuCysTYrSerArg 1295  
DB 213829 CTGCCACCAAGGCTTCACAGGCGACTCTGCCAGACTGAGTGGACTCTGTTACTCAAG 213770  
QY 1295 gProCysGlyProHisGlyArgCysArgSerArgGlyGlyTYrThrCysLeuCysArg 1315  
DB 213769 ACCTTGATGACCCCATGGGCACTGCCGACGTGAGAGGGGTGCTATACCTGCTGTGTCG 213710  
QY 1315 gAspGlyTYrThr----- 1319  
DB 213709 CGATGGCTACACGGGTGAGTCTAGGCGAGGACAAATGGGGGACGCCCTGACAGGCTGACG 213650  
QY 1319 ----- 1319  
DB 213649 CTCTGGGCCAAGCACAATCAGACTAATATAGGGCTGCTCATTTACTTCAGAGGTATA 213590  
QY 1319 ----- 1319  
DB 213589 AACTAGATGCAACAGCGTTCTTAGATCTTACTGCTGTTTGAACAATCGCTGCTTAGTA 213530  
QY 1319 ----- 1319  
DB 213529 AGTCAGAAATAGGAACCTTATGTTCACGTGCAAGCCCAACCACTGSCACATGTC 213470  
QY 1319 ----- 1319  
DB 213469 CGCGTCGCAAGAGCCCAAGAAAGCTGTGAATTGAGTCAGTCCGGGCACTGGGTTGT 213410  
QY 1319 ----- 1319  
DB 213409 GTTCAGCTTTTACCACTAATCTGTGAATAAGAGCAACTTAGCTTCTTCCACAG 213350  
QY 1319 ----- 1319  
DB 213349 AGTATTTAAGATGAGATACCAATTATGTAGTCAAAAAGTTGCTTGTGAGACTCATG 213290  
QY 1319 ----- 1319  
DB 213289 TAAATATTGATAGGAGAGAAAGCTGGCTTATGTGTGAGTGTGGCTTTTGTGAG 213230  
QY 1319 ----- 1319  
DB 213229 GAAGCTCGAGATGAGTGGCTTATCAATTAAAGGTGTGCTGGCTTAGAACAGAGC 213170  
QY 1319 ----- 1319  
DB 213169 CTTCAATATGCTGTGGAATCCCTTCAATTACTTGTAGCAAGACTTCCCAACCAAGAA 213110  
QY 1319 ----- 1319  
DB 213109 CTGAGAGAAAGCACTGAGAAAGAGGGTTTGAGTCAATGGAGTTCCAGAGACACTGAGCC 213050  
QY 1319 ----- 1319  
DB 213049 CGGTAGCCCTGCAAAATGGCTCATTTCTACCTCCAAATGGTCTGGGGCTGAGAGTCT 212990  
QY 1319 ----- 1319  
DB 212989 CGTTAACCCGTGCCCAACCTGTAGGCTGTGCTTCTCTCAGGCTCTTGTGCTGTTC 212930

QY 1319 ----- 1319  
DB 212929 TTCCAGACTCCAGGCTTCAAGTCTTCAGCGCCAGGCGCACTGGCTGSCAGTGGCG 212870  
QY 1319 ----- 1319  
DB 212869 TGGCAGACTCAGTGCCTCTGCTGACACCTCTGTCTAGAACTTGGCAGACTTACTCT 212810  
QY 1319 ----- 1319  
DB 212809 TCTGTGCTGCTGTTGTGAGGGTGGGGGTAGGGGAGGGGGAGTCTTTGTCC 212750  
QY 1319 ----- 1319  
DB 212749 AGAGAGATGGGAAAGGGGCTCCGAAAGTAGCGGAATTAAGTTTCATGGAACAAGA 212690  
QY 1319 ----- 1319  
DB 212689 GCCAGTCACTAGGCAAGGAACTAGAAATCTATCTTAGAATCAGAAATCGGAGGAGA 212630  
QY 1319 ----- 1319  
DB 212629 AGGAGTGGATTTTCCGTCTAATATCTGAATTTTCAAGCCCTGATTTCTGTCTGTC 212570  
QY 1319 ----- 1319  
DB 212569 TCAGGCTTAGGCCCCCAGTCTTCAGGCTTGTTAGGCGGAGGAAATCCTCAGTTGAGA 212510  
QY 1319 ----- 1319  
DB 212509 CCCAATGTTCTGTTCTTGGGGAAGCCCTGCTTACTTACACCATGCAACCCAACTAT 212450  
QY 1319 ----- 1319  
DB 212449 AACTCAATATAGCTCGAAACTTAGTCAATCTTAGGAAAGAAATCTGATCTTGACC 212390  
QY 1319 ----- 1319  
DB 212389 TGACAGAACAAAGCAGAGTGGGAAAGGGGAAAGAACGGAGTGGGGCGCGGCTG 212330  
QY 1319 ----- 1319  
DB 212329 GAACCTGCAAGGGCCAGGGCTTAGGGAACAGGTGACTGACAGGTGTGGGAAAGGCC 212270  
QY 1319 ----- 1319  
DB 212269 AAGAGCTCACAGCAGAACATTTGTGAGCTGACAGAGAAATAGTGTGSCAGGCCAATAAG 212210  
QY 1319 ----- 1319  
DB 212209 AGAGTGTCTGTGCATTAATTTTCAAGATGTCAACCAAGAGAAAGTGAACGATGGGA 212150  
QY 1319 ----- 1319  
DB 212149 GCCCGAGATGAGAGAGGTCAAGAGATGTTAACAGGGAACAGTTAGACTTCTACT 212090  
QY 1319 ----- 1319  
DB 212089 AACTGATTACCCAGAGAGACAGTTTAAACAAGAAAGTTTACTTGTCTCCGGTCTC 212030  
QY 1319 ----- 1319  
DB 212029 TGCAATTCAGTTCTGTGTGCTGTGCTCTTCGCTAATAGACAGACATCAGGGTGGCGAAG 211970  
QY 1319 ----- 1319  
DB 211969 CTGCTACCAAGATGAGAAAGACCTTGGGACAAAGATGAGCCCTGTATGGCGTACTCCA 211910  
QY 1319 ----- 1319  
DB 211909 GTGACTGTCTTCTCAGGCTCTCTCTCCGATGCTTCAGTGTCTCATTAACAGCCCC 211850  
QY 1319 ----- 1319

```

Db 211849 TCAGCAGACCAAGCCTTAGTATATGAGCTGGGGGAGGGGGGCAATTATTTCAAC 211790
QY 1319 ----- 1319
Db 211789 TGTAAACAGGGGACCTTCCAGCATCATCTTGCAGGGCTCACCTGCTTCCCTCC 211730
QY 1320 -GlyGluHisGlyValIleSerAlaArgSerGlyValArgSerProGlyValCysLysAs 1339
Db 211729 AGGGAGACACTGTGAAGTGAAGTGGCCGCTCAGGGCCGTTGTATCCAGAGATCTGCAGAA 211670
QY 1339 nglyGlyThrCysValIleLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAs 1359
Db 211669 TGGGGTACCTGTGTCAACCTGTGGTGGAGGCTTCAATGTGACTGCCATCCGGGGA 211610
QY 1359 PheGluLysProGlyCysGlnValThrThrArgSerPheProAlaHisSerPheIleTh 1379
Db 211609 CTTTGAAGAGCCCTTCTGCAGAGTGAACAAGCAGCTTCCAGCCGGCTCTTCATAC 211550
QY 1379 rPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSer 1394
Db 211549 CTTCCGGGGGCTTGGCCAGGCCCTTCACTTCACTTGCCTTCTGTAGTGAAGAGCC 211490
QY 1394 ----- 1394
Db 211489 CTGGGGTTCGAGGGCACTTGTAGAGAGCTGGGGTACTCTGTGTCAAGTAGACGGGGTG 211430
QY 1394 ----- 1394
Db 211429 CAGATCAGTAAAGAAAGTACCAAGAGAGTGGCTGACAGAGCACTTGATGCTTGCCA 211370
QY 1394 ----- 1394
Db 211369 GCCGGCGCGCTCACTCTCAGAGATCTGGCTGCACTCTCCCATCTTGAGATGGGCG 211310
QY 1394 ----- 1394
Db 211309 GCCAGAAATGCTGGGGTCTGTGATCTGACTCCAGCCTTAGCCACAGTGTCAAGTGTGC 211250
QY 1394 ----- 1394
Db 211249 TGTGCTTAGATGCGCCAGAGTAGTACACGCGGGTGGGTGCCCGCTTCTGCGCCCGTT 211190
QY 1395 ----- 1395
Db 211189 GACCATCCCTTATACAGGTTTGTCTACCAAGAGAGGTGAAGGGCTACTGCTGTAACAAG 211130
QY 1408 PheAlaThrLysGlyLysAspGlyLeuLeuLeuValLeuGln 1408
Db 1408 YArgPheAsnGlyLysHisAspPheValAlaLeuGlnValIleGlnGlnValGlnLe 1428
Db 211129 GCGCTTCATAGAAAGCATGTACTGTAGCTCTCAGAGTGAATCCAGAGCAGGTGCAAGCT 211070
QY 1428 uThrPheSerAla 1432
Db 211069 CACCTTCTCTGCAAGTAGAGCCAGCGCTCTGTCTTGGCCCATCTTAAGATCTTATG 211010
QY 1432 ----- 1432
Db 211009 TCTCCAACTCTGACTCAAGCATGCGCTCCCTCCAGTCAAGACTGAGCTGGGCT 210950
QY 1432 ----- 1432
Db 210949 CCAATCTAGACATGCCCACTTCCCTGCACTCCACTCTTCTGTGACCCAGCTTAC 210890
QY 1433 ----- 1433
Db 210889 CCGTCTTTTCTTAGGGGAATCCACACACGAGTGTCTTCACTTGTGCTTGAAGGGTCA 210830
QY 1448 eArgPheGlyGlnThrPheIleThrValGlnLeuLysIleValAsnLys 1462
Db 210829 GTGATGGCAATGACATACATACAGCTGAAGTACTCAATATAGTGGGTATGGGGGGA 210770
QY 1462 ----- 1462

```

---

```

Db 210769 CAGAGAGCTAGAGGGTCTGTCCGTGGCTGTCTATTGTGAGTGTGGGAAGCTGTTC 210710
QY 1462 ----- 1462
Db 210709 CAGCGCTGGGAAGAGAGAGAGTGTGCTTGCAGGGCACTTCCATCCCTGTGTCT 210650
QY 1462 ----- 1462
Db 210649 TGAATGGGAATATATGTGTAAGCGAGTGTGCTCCGTGGGTGTCTACTGCTTCTT 210590
QY 1463 ----- 1463
Db 210589 CTCTGTCCACAGCGCTGTGTGGGTCAAGACAGACTTCCACAGAGCCCATCTGAGCAGAG 210530
QY 1479 ValAlaValAlaThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerVal 1498
Db 210529 GTGCTGTGGTGTCCGGTGAAGTGTGAACAAGGGGTGAGCTGTGCGCTTGGAGCTATG 210470
QY 1499 LeuGlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyLysSerLys 1515
Db 210469 CTGGGCACTACTCTGTGTGCTGCCAGGGCACCAGAGAGGCAAGAAATGAAGCAGAG 210410
QY 1515 ----- 1515
Db 210409 ATGGGCGGGGAGGGGGAGGGTGCAGAGCTAGAGAGAGGTATGTGGTCAACAG 210350
QY 1515 ----- 1515
Db 210349 ATGGGGCTCCAGGTCAACAGAGATGGGGCTCCGAGTGAAGAACTACATGCAATGCCCTT 210290
QY 1515 ----- 1515
Db 210289 CCTCAAGATTTGGGAATTAGTCAAGTCTTGTGGGCTGTGAAGGGGAAGGGAGCATC 210230
QY 1515 ----- 1515
Db 210229 TGTCACTAGGGGTGCTGTCTCTCGCGCTTTTCTMAAAGGCTCATGAGCTCG 210170
QY 1515 ----- 1515
Db 210169 GCTCAAGTCTTCTTCTTGTGCTTGTAGTGGCAGGGTATGGGTAAAGAAACCAAG 210110
QY 1516 ----- 1516
Db 210109 ACTTAATTCCTGTCCGTGCTTCTTCAATCTTGTGCTTGTGCCAGGCTCTGAGCTGAC 210050
QY 1520 rGlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMetArg 1540
Db 210049 AGGAGCCCTGTGTGCTGGGTGGGTGCCAGATCTGCCGAGAGTTTCCCTGTCCGAATGCG 209990
QY 1540 gGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAs 1560
Db 209989 GCACTTGTGGGTGTGATGAAGAACTTCCAGGTGATAGCCGCACTGCATGGCCGA 209930
QY 1560 PheIleAlaAsnAsnGlyThrValProGly 1570
Db 209929 CTTCATGCGCAACAAATGAGCACTGTGCTGG-TATGAGAACCTGGGGGTACGGCAAGCT 209871
QY 1570 ----- 1570
Db 209870 GGGAGCCAAATGCCATGAGAGTGGTGTGATGACACGGGAATGCCAGCTGCGGTT 209811
QY 1570 ----- 1570
Db 209810 CTTCCCGTCACTGTGTGTTTCCAGAGCCCTGGGGTCCCTGAGTCCCTTCTCCGT 209751
QY 1571 ----- 1571
Db 209750 TCTAGGCTGCCCAACAGAGAAAGTGTGTGACAACTTGGCAATATATGTGGCAC 209691
QY 1588 rCysValAsnGlnThrAspAlaPheSerCysGlyCysProLeuGlyPheGlyLysIle 1608
Db 209690 CTGTGTGAACCAAGTGGAGCGCATTTCAAGTGTGCGAGTGTCCCTTAGGCTTGGGGGGCAAGAG 209631

```

```

QY      1608 rCyba1aGln----- 1611
Db      209630 CTGTGCCCAGGGATGAGAGTGGCAGCTGCCACAGATTGGGGCCCTGGGCTCTGTCAAGTTGTC 209571
QY      1611 ----- 1611
Db      209570 TGGGATTACGGGACAGGGCTTGGGGCAGGCTCTGGGAGGGCTCCGGAGTCAAGCTGTGG 209511
QY      1611 ----- 1611
Db      209510 GTGGAAAGACGCTGTCTGGGACAGTAAAGAAAGTAAAGAGAGAGCAAAAGACCTAAGTC 209451
QY      1611 ----- 1611
Db      209450 AGAGGACAGCCCTGGGACAAATGTAGGGGGGCTGGAAACAAGTAAATGCTCCGGTGGGC 209391
QY      1611 ----- 1611
Db      209390 GGGGACATAGCCAGAGACCCGGGTGTAGGGAGGCGAGGGTCCATCTCATGAGAGCTGCCA 209331
QY      1612 ----- GluMetA 1614
Db      209330 CGCTTCTGACTCCACTGCCCCCTGGCCTTGAATCATCTTTTCATCTGCTTCCAGAAATGG 209271
QY      1614 1AAsnProGlnHisPheLeuGlySerSerLeuValAlaTPrHisGlyLeuSerLeuProI 1634
Db      209270 CCAATCCCAAGGGGTTTCTTGGGACAGGCTTGTGGCTGGGATGGCTCTCTCTGCCCCA 209211
QY      1634 1SerGlnProTPrpYrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuL 1654
Db      209210 TCTCTCAAGCCCTGGACCTCAAGCTCATATGTCGACAGCCGAGGACAGATGGCGTCTTCC 209151
QY      1654 euGlnAlaIleThrArgGlyArgSerThrIleThrLeu----- 1666
Db      209150 TCGAGGCTCTCAACAGGGGGCGAGACCATCACCTTGCAGGTGATGCAGGGCGCTGGGG 209091
QY      1666 ----- 1666
Db      209090 AGTGTGGCAAGCCCCGGCTTGGCTAACTTTGTAGAGAGCTTCAGGGGGTTGGACCCAGGT 209031
QY      1666 ----- 1666
Db      209030 GGCTGGGACAGGGTTGGCAGAAAGCAAGCATGGCTGGGCCCATTCGGTGGTCAAGTACT 208971
QY      1666 ----- 1666
Db      208970 GCCACCTATTGGGCTTAAAGAAATTAATCCCTCTGCTGCTCAGCGCGGACAGCTGG 208911
QY      1667 ----- GlnLeuArgGlnG 1671
Db      208910 AAGAGCCTACTGAGTGTGTCTACTAATTTTCTTCCCTCTCGTACAGCTTCGGGGCAG 208851
QY      1671 1HisIleMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnP 1691
Db      208850 GCCACGTAGTGTAAAGTGTGAGAGGAGGAGGAGGCTTCAGAGGCTCATCCCTGGCTGGAGC 208791
QY      1691 rGcIyArgAlaAsnAspGlyAspTPrHisIleAlaGlnLeuAlaLeuGlyAlaSerGlyG 1711
Db      208790 CAGGCGGACGCATATGATGTGATCTGGCATCAAGCAGCTGTCACTGGAGAGCTAAGGGGG 208731
QY      1711 1PProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGlyIleAsnLeuG 1731
Db      208730 GCCCTGGCCACGCGCATCTGTCTCTTGTGACTATGGGACAAAGAAAGGAGAGGATCTGG 208671
QY      1731 1PProArgLeuHisGlyLeuHisLeuSerPasnIleThrValGlyIleProGlyProA 1751
Db      208670 GCCCTGGCTGACAGGGGCTGCACCTTAGACAAATATGACAGTGGGGAGATTCCCGGGACAG 208611
QY      1751 1aGlyValAlaIleArgGlyPheArgGlyCyS----- 1761
Db      208610 CCAAGCAGTGTGGCCGTGTCTCCGGGGCTGTTTGCAGAGTGAATCTCTCCCCGACCTT 208551

```

---

```

QY      1762 ----- LeuG 1763
Db      208550 CCATCCACTTCTACTGAGAGCCCTGTTCCACTGAGGGCTAAATGCTGCTTATCCCTTC 208491
QY      1763 1ngIyValArgValSerAspThrProGlnGlyVal1AAsnSerLeuAspProSerHisGlyG 1783
Db      208490 AGGCTGTAGGGGTAAAGCAGACACCCGAGGGGTGTAGACAGTCTGGATCCAGCCGCTGGGG 208431
QY      1783 1uSerIleAsnValGlnGlnGlySerSerLeuProAspProCyAspSerAsnProCySp 1803
Db      208430 AAGAGCATATGTGAGAGCCAGGCTGTAGCTGGCCAGATCCCTGTGACTCGAATCCATGTC 208371
QY      1803 rGcIaAsnSerTyrCySAspAsnAspTPrAspSerTyrSerCySAsp----- 1820
Db      208370 CCACCAACAGCTACTGACAGCAACGACTGGGACAGCTATTTCTTGTATGCTGTATCCAGSTA 208311
QY      1820 ----- 1820
Db      208310 AGCTGAGAGTCTGGGAAATGGGGCTGGTGCCACGGGTCACTCCGTCTTAATATGTA 208251
QY      1820 ----- 1820
Db      208250 GCCAGTTACATCCAGGGTATAGCCATCTGGTGCAGGGTTCAAGCGGAGTAATAAGGA 208191
QY      1820 ----- 1820
Db      208190 GTTGTATAGTGAAGAAAGATGCCAGTAAAGGGCGGGTCTGAACCTTTCGATCCTTCTT 208131
QY      1821 --ProGlyTyrTyrGly1AAspAsnCySThrAsnValCySAspLeuAsnProCySgIuHisG 1840
Db      208130 GTCCAGGTACTTATGTGTAACTAAGTAACTGTAACAAATGTGTGTGAACCTTAACCAAGCAGACC 208071
QY      1840 1uSerValCySThrArgGlySerProSerAlaProHisGlyTyrThrCySgIuCySProProA 1860
Db      208070 AGCTGTGTGTATCCGAAACCCAGTGCACCCCAAGGCTATCATCTGCAGTGTTCACCA 208011
QY      1860 snTyrLeuGlyProTyrCySgIuThrArgIle----- 1870
Db      208010 ATTACTTGGGCCCCATATGTGAGACAGGTA--GGTGGCGGGGCTTACTCTGTATGAC 207952
QY      1870 ----- 1870
Db      207951 TCAAGCCCGAAGTGCAGAGTGGGCTGAGCTGCAGCCCATGAAAGACTAATGTGGT 207892
QY      1870 ----- 1870
Db      207891 GAACCAAGCGGAGAGATGCAGTGTAGAGGGGCTCATAGCTGTTGGGTCCTTCT 207832
QY      1871 ----- AspGlnProCyAspProArgGlyTyrTPrGlyHisProThrCySgIyProCySA 1888
Db      207831 GCAGGATTTGACCAACCTTGGCCCCGCTGGCTGTGTGGGACACCCCAATGTGTCCATGCA 207772
QY      1888 snCyAspValSerIyGlyPheAspProAspCyAsnIleThrSerGlyGlnCySHisC 1908
Db      207771 ACTGCAATGTGACAAAGGCTTTTACCCAGACTGCACAAACAAACAAAGTGGCAGATGCCACT 207712
QY      1908 ySlyS----- 1909
Db      207711 GGAAGTATAGAGACCCGAATGAGCCCCGCTGTGTACTCCGTCTGTAGCTTGGGCC 207652
QY      1909 ----- 1909
Db      207651 TACAGGGAGACAGAGAGATGGGCTTCTCTCAGAGAGCGAGACAGCATCTGTGCTT 207592
QY      1909 ----- 1909
Db      207591 TTGCTTATCAGGCAATGGGAAGTGTGTTTGTGCTTATATCCGCTGCAGGGGAGATCCA 207532
QY      1909 ----- 1909
Db      207531 GTTCCAGTGTGAGTCAAGTGAAGCTGACATAATGTGTGGGCTGTTTGCAGGCCATTTCTG 207472
QY      1909 ----- 1909

```



Db 207471 AGTATTTGGCTTTCACCTGTAATGTCATTACCTGCTGTGAGAGTGTGAATCTCTCC 207412  
QY 1909 ----- 1909  
Db 207411 CACAAGACCCAGCATAGGGAGGCTCTCTCTGGGGAGGCCACAGCTTGAAGCTTCGCT 207352  
QY 1910 ----- GluAsnHisT 1913  
Db 207351 AGCATCCAGAAAGACATCTTGATTCAGTGAACCCCTTCTCCCTCCCTGAGAGAACTACT 207292  
QY 1913 YTAAGProGlySerProThrCysLeuLeuCyAspCysIleProThrIleSerLeus 1933  
Db 207391 ACCGACCCCGAGCGAGCCCACTGCTCTGTGTGACTGTACCCCACTGGCTCTTGT 207232  
QY 1933 eATGValCyAspProGlyuAspGlyGlnCysProCysLeuProGlyValIleGlyAarg 1953  
Db 207231 CCGAGTCTGAGCCCTGAGAGAGCCAGTGTCCATGCAAGCTGAGTCAATGGGCGTC 207172  
QY 1953 InCyAspArgCyAspAsnProPheAlaGluValThrThrAsnGlyCyAsGlu----- 1970  
Db 207171 AGTGTATCGCTGTGACAAACCTTTTGTGAGGTCAACCAATGCTGTGAAGGTAGG 207112  
QY 1970 ----- 1970  
Db 207111 GGCTCTTAGCAGAGTAGTTCCTCCTCACAATATGCAAGAGCTTACACTCGAAG 207052  
QY 1970 ----- 1970  
Db 207051 TTAGAAATGAGGCTGGCATTTGGGTGGGAGATGTGAAGAAAGTTGGGGGATGATAG 206992  
QY 1971 ----- ValAsnTyrAspSerCyAsProArg 1979  
Db 206991 AACTGCTTACCGATACCATATGTTCTTCTCCATATGAAATTATGAGAGCTGCCACGG 206932  
QY 1979 laIleGluAlaGlyIleTTrProArgThrArgPheGlyLeuProAlaIleAProC 1999  
Db 206931 CCATAGAGGCTGGAGTCTGTGGGCGCCGACGCGGTTGGGCTACTGCTGCTCCCT 206872  
QY 1999 YsProLysGlySerPhe----- 2004  
Db 206871 GCGCCAAAGGGCTCTTTGTGTAGTTTGAATTACATATGATTTGAACACTTGACTTT 206812  
QY 2004 ----- 2004  
Db 206811 ACTCTCAGAGGAATAAAGATAGACAGTACCGAGCTTGAAGCTCCGGGGGCCCACTT 206752  
QY 2004 ----- 2004  
Db 206751 TCTGGATTCTATCTCCAGAGACTAGTGTCTCACTTCACTGACTCAACCATCAA 206692  
QY 2004 ----- 2004  
Db 206691 CTTAGGCTGACAAACGTGCTTATATCCAGCCCAACAGAGCTGACGAAATCTGCGTG 206632  
QY 2004 ----- 2004  
Db 206631 TCTGAAACAGGGCTGGAGCTAGACTGTGTGATTGCAATGGGGTGCCTATCTTGAGCTAAC 206572  
QY 2004 ----- 2004  
Db 206571 TACTTAGTACTGAGTTACCCCACTGAAAGGAGCGGTGCTGAGAGTATGCGTGT 206512  
QY 2004 ----- 2004  
Db 206511 GGGTACTATCCCAAGAGATTCAAGTATAGTGTGACATGTTTCAATGGGAGGAGTAGCT 206452  
QY 2004 ----- 2004  
Db 206451 GAGAGTAAGCAGTAGCTGATGACAGGCGTAGTATCTTGGGGCGGTGTGGAGA 206392  
QY 2004 ----- 2004

Db 206391 GGTACATAGGCTCCTTTGGCCACTGACTGATTTTGTGAGATGTTTGGCTCTGGAG 206332  
QY 2004 ----- 2004  
Db 206331 CTCTGACTGGGCTTCTCTGTCTCCTAAGCCCGCTTCTTGCTTCTTCTCA 206272  
QY 2005 GlYThrAlaValArgHisCyAspGluHisArgGlyTTrLeuProAsnLeuPheAsn 2024  
Db 206271 GGAATGCTGTGGCCCACTGTATGACACAGGGGCTGGCTCCGCAAACTTTCAC 206212  
QY 2025 CysThrSerIleThrPheSerGluLeuLysGlyPheAla----- 2037  
Db 206211 TGCACTCACTCACTTCTCAGAGCTTAAAGGCTTTGT-AAAGTTAGCCCTTGTCTTCTC 206153  
QY 2037 ----- 2037  
Db 206152 CTCTTCTTATACCTCTGCTCTTGAATAATAGACTCCGGGGTTGGGATTTAGCTCAGT 206093  
QY 2037 ----- 2037  
Db 206092 GGTAGAGCGCTTGCTTAGAAGCGCAAGGCCCTGGTTCGCTCCAGCTCCGGGAAAAA 206033  
QY 2037 ----- 2037  
Db 206032 AGAACCAAAAAAAAAAAAAAAAAAATAGAGCCGAAATCTGTTTCTGAGGCTCT 205973  
QY 2037 ----- 2037  
Db 205972 CTTACTGACAGAGCGCATTTCTTCCCACTGATCTTTAGGCCCTTTTACTGTGAGCC 205913  
QY 2038 ----- GluArg 2039  
Db 205912 CAGTCCCAACCATCTGTGTGACCTCTGCACTTGTGCTCTCCATCCAGGCCAGCGG 205853  
QY 2040 LeuGlnArgAsnGlyuSerGlyLeuAspSerGlyArgSerGlnGluLeuAlaLeuLeu 2059  
Db 205852 CTACAGCGAAACCAATCAGGCTGTGACTCAGAGCGCTCCAGAGGCTGGCCCTGCTCTG 205793  
QY 2060 ArgAsnAlaThrGlnHisThrAlaGlyTTrPheGlySerAspValLysValAlaTyrGln 2079  
Db 205792 GGTATGCCACACAGCACCTCTGCTACTTGGGCACTATGTCAAGGTGGCTTACAG 205733  
QY 2080 LeuAlaThrArgLeuLeuAlaHisGlyuSerThrGlnArgGlyPheGlyLeuSerAlaThr 2099  
Db 205732 CTGGCCACAGACTGTGCTGATGAGAGTGCACAGGGGCTTTGGGCTGTCCGCACA 205673  
QY 2100 GluAspValHisPheThrGlu----- 2106  
Db 205672 CAGGATGTGACTTCACTGAGGTGGGGCTTGAAGTAGCATGTGGGGCCAGTGAGAGA 205613  
QY 2106 ----- 2106  
Db 205612 GCTGTTCCTGAGATCTTGGGGGAGATCTTGAAGGAGATCTTCTTACTGTAAAGAGC 205553  
QY 2106 ----- 2106  
Db 205552 TGCCCAACCCCTGTGACAGAGGCTCCATGATAGTCTGTAAACAGAGCGCAACAGTTG 205493  
QY 2106 ----- 2106  
Db 205492 ATTAGAGTACTCTCTCTGGGCTGGGATTTAGCTCAGTGTAGAGCGCTTAACTAGG 205433  
QY 2106 ----- 2106  
Db 205432 AAGCGAAGGCTGTGGTTCGTGCCAGCTCCGAAAAAAGAAAAAAGAAAGTA 205373  
QY 2106 ----- 2106  
Db 205372 CTCTCTTACAGTGTCAAGAAATTAGTACTAGCTTGTGATGAGGGCTGACCTGAGGCA 205313  
QY 2106 ----- 2106  
Db 205312 GAGCCCACTTATTAACAAGAGTCACTTGTGTGAGGAGCCAAAGAGGCAAAAAACA 205253

QY 2106 ----- 2106  
 Db 205252 GAGCGACATGCAAGACACATGGGTGGCCCTGATTTAAAGTGAAGAGGTTTAA 205193  
 QY 2106 ----- 2106  
 Db 205192 GGCAGGATCCATGCTGGAAGTGAAGATATGATACAGAGGCTTCCAGGGGAGACTG 205133  
 QY 2106 ----- 2106  
 Db 205132 TGGGACTGCTAAGGAGGAGACCTAGTGGCGGGGTCTGAACCAAGGCAAGGCTGAACTT 205073  
 QY 2107 ----- 2109  
 Db 205072 GGTATGAGGTTGACCAAGCATATGCACTGTGACCATGCCCATCTCCAGATCTGCTG 205013  
 QY 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisTyrGluLeuIleGln 2129  
 Db 205012 AGGGTGGGAGCGCCCTCTGGATGCGGCAATTAAGGCACTGGGACCTGATCCAGCAG 204953  
 QY 2130 ThrGluGlyThrAlaTyrLeuLeuGlnHisTyrGluValTyrAlaSerAlaLeuAla 2149  
 Db 204952 AACAGAGGTGGCACTGCTGCTGCTCCAGCACTAAGGCTTATGCCAGTGGCTGGCC 204893  
 QY 2150 GlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThrProAsnIle 2167  
 Db 204892 CAGAAATATCGCGCATACTACTAGAGCCCTTCACTATCGTCAACCAACATTTGGTAG 204833  
 QY 2167 ----- 2167  
 Db 204832 GTGGTCTGGCGTGGGGGCGGGGGTGTGGAGAGTCTGCCCATTTCAAGGGCAGCCAG 204773  
 QY 2168 ----- 2174  
 Db 204772 CGGTGCTGTGTAGTACTTTCGCAACCAATTTCTTCCTAGTCACTCTGTGGTGGCCCTG 204713  
 QY 2175 AspLysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgLysGln 2194  
 Db 204712 GATTAAGGGGAACCTTCTCTGGGACCAAGCTGCCCGTTATGAGGCGTTAGTGGGGAGCGC 204653  
 QY 2195 ProProAspLeuGluThrThrValIleLeuProGluSerValPheArg 2210  
 Db 204652 CTTCCAGATCTTGAGACCAAGCTATTTTGGCCAGAGTGTCTTCCAGAGGTCAAGCGGGGA 204593  
 QY 2210 ----- 2210  
 Db 204592 AGTGTGATGTGTAACGGCTTGGGGGTGGGGCAACCATGCAAGTGGGAGGGCCAGCTGAT 204533  
 QY 2211 ----- 2224  
 Db 204532 GTAGCAGCATTTACTTGCAGAGATGCTCCCATGTGAGATCTGAGAGACTGCTGTAAAC 204473  
 QY 2224 agInguLysProGluLeuAlaArgArgGlnArgGlyAspProLysLeuSerGlnIle 2244  
 Db 204472 CCAAGGACTGAGAGACTGGCCCGGGCGGAGGAGGCCCAAGAACTAAGTCAAGGGCCA 204413  
 QY 2244 uAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAs 2264  
 Db 204412 GGCAGTGGCCAGTGTCACTATTTAACACACGCTGCTGAGTGTGCTGCCCAACACTAGA 204353  
 QY 2264 pProAspLysArgSerLeu 2270  
 Db 204352 CCAAGATTAAGCCGTAGCTGAGGTAAAGCGGCTACGACAGGCTTAAAGAGTTGGGGTTAAC 204293  
 QY 2270 ----- 2270  
 Db 204292 AAGGAAATGTGAAGATCTCCCAAGTATTAAGGTTCTCTGCAAGCGGCTTTGGTACTGT 204233  
 QY 2271 ----- 2285  
 Db 204232 GTACCTTCTGTATCATAGAGTCCCAAGGCGCCAGTCAACACACTGTGTGTAAAGCAT 204173

QY 2285 eSerValHisAspAspGluLeuLeuProArgAlaLeuAspLysProValThrValGlu 2305  
 Db 204112 CAGTGTCCAGATGATGAGAGGCTCTCTCCACGGGCACTGGACCAAGCAGTCAAGTGA 204113  
 QY 2305 nPheArgLeuLeuGluThrGluLysGlnThrLysProIleCysValPheTyrAsnHisSe 2325  
 Db 204112 GTTCCGACTGTGAGAGCGGAGGAGCGAACCAAGCCCATGTGTCTTCTGAAACATATTC 204053  
 QY 2325 rIleLeu 2327  
 Db 204052 GATTCGTGAGCTGCTGTATCCACCTTTAAGCCCATGACTTTGGACCAAGAGTCCAGGC 203993  
 QY 2328 ----- 2331  
 Db 203992 TCCTGGCTGTCACTCTTCTTCTCCCTGACTCCGCTTCTCTCCAGAGGCTCAAGTGA 203933  
 QY 2331 hrgLysIleTyrSerAlaArgLysGlyValValPheArgAsnGluSerHisValSerC 2351  
 Db 203932 CAGGTGGCTGTGCGCCAGAGGCTGCGAGGTGTCTTCCGTAAAGAGGCAATGTCAAGCT 203873  
 QY 2351 ySGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArgArgLys 2369  
 Db 203872 GCGAGTGCACCATATGACAAAGCTTTGCTGTCTCATGAGATGTCTCCAGACAAAGCTTG 203813  
 QY 2369 ----- 2369  
 Db 203812 GACTCCAGGGGTAGCCCAAGAGCAGCGGGGTGGGACCAAGTCAAGGGCTCAAGGCTTG 203753  
 QY 2370 ----- 2379  
 Db 203752 CTCTTTTTCACCCCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 203693  
 QY 2379 eutThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeu 2399  
 Db 203692 TGAGGTATGTGGCCCTTGGAGTCACTTACTGTGCTGTATGATACCTTCTCTTCCCA 203633  
 QY 2399 hIleuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAla 2419  
 Db 203632 CCTCTCTTCGAGCCCTTCTGCTCCCAACGAGTGCATCCGACGCAACCTCAAGCTGCCC 203573  
 QY 2419 euGlyLeuAlaGluLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuPro 2436  
 Db 203572 TGGGCTGGCCAGCTGTCTTCTCTGGGAAATCAACAGGCTGAGCTT - CCTGTAAAGA 203514  
 QY 2436 ----- 2436  
 Db 203513 TACTTCTCTCCAGAGACTTCCCAACTTCAAGCCCAACCAAGGCTCATGTGTCCCT 203454  
 QY 2436 ----- 2436  
 Db 203453 CTATGACCTTGTCCATCCCTGCTCCACAGCAATCAGTAGTAAAGGCCCATTTCCAGT 203394  
 QY 2436 ----- 2436  
 Db 203393 ACTGCGCCAGATTGGCCCTTCTCTCAACACACAGGCTTCCCACTCCCTGGGGGAGC 203334  
 QY 2437 ----- 2441  
 Db 203333 TCCTTGTATTAAGTGTGGCTCTGTATCATATGTTCCGTGGCCCTCAGTTTGTTCGACAG 203274  
 QY 2441 aIleIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGlu 2461  
 Db 203273 TCATTTGCATCTGTCTGACATTCCTGTACCTTTGACACCTTCTCTGGGCTCTGTGTGAAG 203214  
 QY 2461 lAlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgP 2481  
 Db 203213 CCTTACACCTGTACCGGGCGGTCAACAGGTGCGAAGCTCAATGCCAGCCCATGTGCTT 203154  
 QY 2481 heTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr 2494  
 Db 203153 TCTACTACATCTGGGCTGGGGCGTTCCTGTCTTCAATCAACAGTATCTCGCTCTTCC 203094  
 QY 2494 ----- 2494

Db 203093 TGAACCTAGAGCTCCCTGTTGCTAGAGCTGAGGGGTACCTTCATGATGATACCCCAAC 203034  
 QY 2494 ----- 2494  
 Db 203033 CATAGAAATTCCTGAGAAATCCCTTCCCTTACTTGGACAGCCTGGATTAGGTGTGAGG 202974  
 QY 2494 ----- 2494  
 Db 202973 GCTAGAGCTTCTCTTGGAGAGGAGTGTGACAGAGGTCTCTGTATGTCTCCCACTG 202914  
 QY 2494 ----- 2494  
 Db 202913 GCTCTGTGAACCTTGAATGATCATCTGCTCAACCTGAGACCTGTCTGTGCT 202854  
 QY 2495 -----GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAsp 2510  
 Db 202853 GTCTCTGTGCTCCAGGTCTGGCTGTGGGTGGATCCGAAAGCTATGAAACCTGACT 202794  
 QY 2510 heCyTrpLeuSerIleTyrAspThrLeuIleTrpSerPheAlaGlyProValAlaPhe 2530  
 Db 202793 TCTGCTGGCTCTTCATCTACATACGCTCATCTGAGTTTGTGTGACCAAGTGGCTTTG 202734  
 QY 2530 lValSer----- 2532  
 Db 202733 CTGTTTCGGTAAAGTCTTGAAGCAATTAGGATGACAGATGCGAGGCTCTACAGAT 202674  
 QY 2532 ----- 2532  
 Db 202673 TCCCTTTAGGAAGCTAGAGCTCTGAGGCCCCCTGGGCAAGAAAGTCTCTGCTCC 202614  
 QY 2533 -----MetSerValPheLeuTyrIleLeuAlaAla 2543  
 Db 202613 TCTCTGCTCTCATATGATCCCACTGACATGATGATGCTTCCGTGATATCTGTGCGGCC 202554  
 QY 2543 hGAlaSerCyAlaIleAlaGlnArgGlnGlyPheGluIleValGlyProVal----- 2559  
 Db 202553 GAGCTCTGCTGCTGCTCCCAACGTCAGGGCTTCGAGAAAGGCCCTGTGTGATGATCTGG 202494  
 QY 2559 ----- 2559  
 Db 202493 CAGGTGGCTGGGCAACAGGCGGGTGTGTATAGAGATCATGGCGGGTGTGTGATAT 202434  
 QY 2560 -----SerGlyLeuGlnProSer 2565  
 Db 202433 GAGAGCATCATGCGGAACTCAATATGCTGCTCCCTCAAGCTCAGGCTGCGCTCTCC 202374  
 QY 2566 PheAlaValLeuLeuLeuSerIleThrTrpLeuLeuAlaLeuLeuSerValIleAsn 2585  
 Db 202373 TTCACTGCTCTCTGCTGTGATGATGACATGCTGCTGGCACTGCTGTCTGTCAACAGC 202314  
 QY 2586 AspThrLeuLeuPheIleTyrIlePheAlaThrCyAsnCyAla----- 2600  
 Db 202313 GACACTCTCTCTTCACTACCTCTTCCGGGCTGTGAATTGTGT-CCAGGTACCTAGCCC 202255  
 QY 2600 ----- 2600  
 Db 202254 CGCCTTACTGAGTGAAGAGATGCTGAGGGGTGGGTGTGTCTCAACCATGAGCTGTCT 202195  
 QY 2601 -----GlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGluValArgIys 2618  
 Db 202194 GCCTGCAAGGGCCCCCTTCAATCTTCTCTCACTAGTGTCTCAAGAGAGGTGCTGGAAA 202135  
 QY 2619 AlaleuValLeuAlaCySerArgIysProSerProAspProAlaLeuThrThrIysSer 2638  
 Db 202134 GCACTCAAGTTGCGCTGAGCGGGAAGCCAGCGCTGACCTGCTGTGACCACTAAGTGT 202075  
 QY 2639 ThrLeuThrSer----- 2642  
 Db 202074 ACCCTGAACCTC-GGTAGAGGGGCGGGGTCTGAGCTCAAGAGCGCTTTGTATGAAGGG 202016  
 QY 2642 ----- 2642

Db 202015 AAGCAGCCCCAGAAACCTGTTGTCTTWTGATCTTTTAACTGGGATCCTTTTCC 201956  
 QY 2642 ----- 2642  
 Db 201955 CATCTCTCATCAGAGAGAGGGGTGAAGTGAAGGGGTGTCTCTTGGAGAAATGCTA 201896  
 QY 2642 ----- 2642  
 Db 201895 TCACGTGGTAAAGCTTTTGACTTGGGGGAGATGCAAGCAAGATTTGGTTGTCAAGTGG 201836  
 QY 2642 ----- 2642  
 Db 201835 ACACATGTGTAGAAATTCCTCATGAATTTGTATTAAGTTTCTGTCTCATAATAT 201776  
 QY 2642 ----- 2642  
 Db 201775 CTCAGAGACATTCCAAGTCCATTAATAAAAAAAAAAAAAAAAAACAGCAAGCCTTATA 201716  
 QY 2642 ----- 2642  
 Db 201715 CAAGTCCGGGGGTTTGTCTGTATTAGTCAACAGCTACGGTATAGCTGTGCCACTG 201656  
 QY 2642 ----- 2642  
 Db 201655 CTGTGCCCTGTGTGAAGCAACAGAGGCGCTTCTCGTTAGCCCTGGGCAATGCGG 201596  
 QY 2642 ----- 2642  
 Db 201595 AGGTAGGGCTTGGGTAAAGTTCCTGAGGCGCACCCCTCCACTCAACCACTCATCTGGCAT 201536  
 QY 2643 -----SerTyrAsnCyAspProSerProTyrAlaAspGlyArgLeuTyr 2656  
 Db 201535 GTCTGTCTCTTGCCTAGTCTCTATTAAGTCCCAAGCCCTTATGAGATGGAGGCTGTAC 201476  
 QY 2657 GlnProTyrGlyAspSerIleAlaGlySerLeuIleSerThrSerArgSerGlyIysSerGln 2676  
 Db 201475 CAGCTTAAAGAGATTCGGCTGCTCTTGTGACAGTGCACGCCCTCAGGCAAGATCAG 201416  
 QY 2677 ProSerTyrIleProPheLeu----- 2684  
 Db 201415 CCAAGCTACATCCCTCTTCTGTCTGATGATGACCCCGGAAATGGAGAGCTGGAAAGAG 201356  
 QY 2684 ----- 2684  
 Db 201355 GAAAGACTGATGCTAGAACCAAGGCGGCGAGCTGTAGAGCGAGTGTCTAGACTG 201296  
 QY 2684 ----- 2684  
 Db 201295 CGGCCAACAGGTCAACAGTTTGAAGCATGAAGGACAAAGGTAGGCGCTGGAAATCT 201236  
 QY 2685 -----ArgGluIleSerAl 2689  
 Db 201235 CACTAGAGAGGCGACATGACCCCTTACAGCATAGTCTTGTCTTTTAAAGGAGAGTCCAC 201176  
 QY 2689 AleuAspProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeuPheLeuGln 2709  
 Db 201175 GCTGAACCTGTGGCAGAGTTCCCTCTGAGGAGACCAAGGTGGCTATTCATGAAAGG 201116  
 QY 2709 yGluAspGlnGlnHis----- 2714  
 Db 201115 CCAAGCCAGCAACAGGTGAGAACAGAGCCTTAGATGCTTACAGGCGAGGGCGCTGG 201056  
 QY 2715 -----A 2715  
 Db 201055 CTTTTCGTGAAGGGGGGTGGGCGAGGTGTGTATTCCTCCTCCTGACCTCCAG 200996  
 QY 2715 spProAspThrAspSerAspSerAspLeuSerIleGluAlaAspGlnSerGlySerTyr 2735  
 Db 200995 ATCCCGACACCGACCTGACAGTGAACCTGTCTCCGTGAAGATGACCAAGATGTCTTACG 200936  
 QY 2735 lAserThrIleSerSerAspSerGluGluGluGluGluGluGluGluGluGluGluGluGlu 2755  
 Db 200935 CTCTACCCACTCATCAGACAGC-----GAGAAAGAGAGAGAGGCGCGCT 200888

Oy	2755	hePrOglYvLuInguLgYTpaSPSeRLeuEnGLYProGluYalagIuaRGLeuProLeuH	2775
Db	200887	TCCCTGGCGAGCAGGGGCTGGACAGCCTGTGGGGTCTTGAGCTGAAGAATGCCCTTC	200828
Oy	2775	iSeSerThrProLyS-----	2779
Db	200827	ACAAGTACCCTCAA-GGGTGGGCAGACAGCACGGGCTGCAAGCATTTGGGGGGGACAAAG	200765
Oy	2779	-----	2779
Db	200768	CCTAAGAGCTTAGGGCTTCTTGCTATCTGTACTCCTCTCTGGAGACTCATAGGCTTCC	200709
Oy	2780	-----AsPglYglYProGlYProGlYLYsAlaProTrPPProGlYAspPhEglYT	2796
Db	200708	TTTTCTTTGACAGATGGGGGTCCAGGGTCTGGAGAGGTCCCTTGGCCAGAGACTTTGGGA	200649
Oy	2796	htThrAlaLySGlUSeRSeRglYaSngLYAlaProGlUGluLaRGLeuARglYaSngLYA	2816
Db	200648	CCACAACAAAGGAGAAATAGTGTAGTGGGGCCCCCTTAAGAGCGGCCAACGGGAATAAGAG	200589
Oy	2816	sPlAlaLeuSeRtARGlUGlYSerLeuGLYProLeuProGlYSeRSerAlaGlNProHIL	2836
Db	200588	ATGCTCTAACTGGGAGAGGGTCTCTGGAGACCCCTTCGGGCTCTTACCACTCACA	200529
Oy	2836	YSGLYIlE-----	2838
Db	200528	AAGGTTA-GTGAAGAGTCTCTCCATCTGCTTCACTCTGCACTCTCTTGCGAAATTTT	200470
Oy	2839	-----LeuLYsLYsCYsl	2844
Db	200469	AGAAATGCTCATCTCATGTAGTACCCACACTCTTACCAGGACATCTCAAGAAAGATGTC	200410
Oy	2844	eUProThrlleSeRglULYSeRSerLeuLeuArgLeuProLeuGLUGlnCYStHglYS	2864
Db	200409	TGCCCACTACGACGAAAAGATGAGCTCTTAAGTTGCCCTCGAGACAGGGCACAAGGT	200350
Oy	2864	eRSeRARglYSeRSerAlaSeRGLUGlYSerARglYglYProPProARglYProP	2884
Db	200349	CTTCCGGGGGCTCCACCGCCATGAGAGGAGCAGCAGAAACGAACTCTTCCACGCCACAC	200290
Oy	2884	rOARglnSeRLeuGLUGlUGlnLeuASngLYALMetProIIeAlaMetSerIIeLYaA	2904
Db	200289	CGCGCCAGCTCTCTCCAGGAACAGCTGAATGGGGTCAATGCCATGCGCATGACATCAAG	200230
Oy	2904	IaGlYThValaSpGIuaSPSeRSeRglYSeRGLU	2915
Db	200229	CAGGCACTGTGATAGAGACTCTTCTGGCTCCGAG	200195
<b>RESULT 13</b>			
AC095463/c			
LOCUS AC095463 235327 bp DNA linear HTG 09-MAY-2003			
DEFINITION Rattus norvegicus clone CH230-7H8, WORKING DRAFT SEQUENCE.			
ACCESSION AC095463			
KEYWORDS AC095463.6 GI:30467711			
SOURCE HTGS PHASB2; HTGS DRAFT; HTGS_FULLTOP.			
ORGANISM Rattus norvegicus (Norway rat)			
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 235327)			
Munzy,D.Merie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,			
Allen,C., Allen,H., Alsbrooks,S., Amth,A., Anguitano,D.,			
Ayalelechit,V., Ayogei,A., Ayodeji,M., Baca,E., Baden,H.,			
Baldwin,D., Banderamake,D., Barber,M., Barnstead,W., Benahmed,F.,			
Bernalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,			
Byrnes,N., Bukey,C., Burch,P., Butrell,K., Calderon,E.,			
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,			
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,			
Cleland,C., Cockrell,R., Cox,C., CoyLe,M., CreA,A., D'Souza,L.,			
Dayila,M.L., Davis,C., Davy-Carroll,L., De Ande,C., Dederich,D.,			
Dejgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,			

Draper, H., Dugan-Rochs, S., Dunn, A., Dublin, K., Dival, B., Eaves, K.,  
Bran, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabler, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunnaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howell, S., Hulys, S., Hume, J., Idelbrd, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C.,  
Kowis, C., Kraft, C.L., Lebrow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshewal, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,  
Mashbawari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,  
Mishra, V., Moriga, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwachilemeh, O., Okunolu, G., Olamunsgoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C.,  
Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Piazo, M., Quiroz, O., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savary, G., Scheer, S., Scott, G., Shetsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taboor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanil, K.,  
Valae, R., Vera, V., Villasaana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
Williams, G., Willson, R., Wlaczky, R., Woodson, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Weinert, G., and Gibbs, R.A.

Unpublished  
Direct Submission  
2 (bases 1 to 235327)  
Worley, K.C.

Unpublished  
Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 235327)  
Rat Genome Sequencing Consortium.

Unpublished  
Direct Submission  
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On May 9, 2003 this sequence version replaced gi:24941094.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a contig-scaffold). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by extended gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GCGO  
Center clone name: CH230-7H8



Qy	541	GIUTYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560	Qy	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProVal	920
Db	223194	GAATACAGCTGGCTGGGGTGGGCAAGACTTCCCTTCAACATTAAACAAGGCAAGGC	223135	Db	222114	GGCCGCAACACCATTGAAGTGAAGTACTGTATCGATGTGAATGACAATCCCGCTC	222055
Qy	561	TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal	580	Qy	921	PheGluGlnAspGluPheAspValPheValGluGluLeuAsnSerProIleGlyLeuAlaVal	940
Db	223134	TGGATTCGGTGGCCCGGAGTGGACCGGGAAGAGTGTGATTCTACAGCTTCGAGTA	223075	Db	222054	TTTGAAACGAGTGAAGTTGATGTATTTGTGAAAGAACACCCCATGTGAGCTGGCTGTG	221995
Qy	581	AlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600	Qy	941	AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle	960
Db	223074	GAAGCAAGGAGCAACCGGCAACCCAGCACTACCGCTCAGCCAGTGTCACTGAACATC	223015	Db	221994	GGCCGGGTCACAGCCACCGGACCCAGACGAAAGGACCAATGACAGATATGACAGTAC	221935
Qy	601	LeuAspValAlaAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620	Qy	961	ValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla	980
Db	223014	CTGAGTGTCAATGACAACAACCACTTTACGACCGGAGTACACGGTGGCGGCTCAAC	222955	Db	221934	GTGAGGGCAATATCCCTGAGGTCTTTCAAGCTGGAATCTTCTGGCGGAGTAAAGCC	221875
Qy	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640	Qy	981	LeuValAspLeuAspTyrGlyLeuAspArgProGluTyrValIleLeuValIleGlnAlaThrSer	1000
Db	222954	GAGGACGGCGGTGGGCAACAGTGTGTGACGGTGTCAAGCCCTGATGAGATGTCTAC	222895	Db	221874	CTGTGATTTGGATGATACGAGGACCGACCTGAATATATCTGTGTATCTAGGCTACGCT	221815
Qy	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660	Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro	1020
Db	222894	AGGTGATTAACCTACAGATCACACGGGCAACCCGCAACCTTCTCTATCACACAC	222835	Db	221814	GCTCCCTGTGGAGGAGGCTATGTCTGTATCCGCTCTCGACCGCAATGATACCTG	221755
Qy	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluArgGln	680	Qy	1021	ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer	1040
Db	222834	CAAAAGGGGGTGGCTGGTCTCCCTGGCTTACCCCTGGAATCAAACTCGAGGGGCAAG	222775	Db	221754	CCAGGTGTGGCAACTTTGAGATCTTTTCAACAACATATGTACCAACCGCTCGAGAGC	221695
Qy	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700	Qy	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	222774	TATGTCTGGCCCTGAGCTGCTGTGATGTGCAAAAGGACAGACACACTCAATCGGTGG	222715	Db	221694	TTCCCTGGGGGTGCTATAGGCGCGGTGCTGCCATGACCCCGAATCTCAGACGCTG	221635
Qy	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720	Qy	1061	ThrTyrSerPheGluArgGlyAsnGlyLeuAsnSerLeuValIleLeuAsnAlaSerThrGly	1080
Db	222714	AATGTCACTGATGCCAACAACCACTGCTCCGCTTCTCAAGCTCCCACTATACGGTGAAT	222655	Db	221634	ACGTACACTTTGAGCGAGGAAATGAACTCAAGCCGTGCTCTACTCAATGCTCCACCGGT	221575
Qy	721	ValAsnGluAspArgProAlaGlyThrThrValIleLeuIleSerAlaThrAspGluAsp	740	Qy	1081	GluLeuTyrLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	222654	GTTAAAGAAAGACCGGCAAGGACGACCAAGTGTCTGATGCTGATCAAGTCAAGAGAGC	222595	Db	221574	GAGCTGAGACTGAGCCGGGCACTGGAACAACCGGCCCTGGAAAGCATCATGTAGTGTG	221515
Qy	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760	Qy	1101	LeuVal	1102
Db	222594	ACAGGGAGAGATGCCGAATCACTTATAGAGAGACAGCATCCCTCAAGTCCGCAATC	222535	Db	221514	CTGGTGTAGGTAAAGAAAGATGCCAGTATGTGGGTGGGAATATGCTTTAGGAGG	221455
Qy	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGlyLeuAspGlnValSer	780	Qy	1102	TCCTTGAAGCACTGGAGGTGTGGCTTCCAAAGAGGGGCTGTGACCAAGCTCTCAC	221395
Db	222534	GATGCAAGACCTGGGGCCGTCACACCCAGGCTGAGCTGAGCTATGAAATCAGGTGTCT	222475	Db	221454	CTTGTGAAGCACTGGAGGTGTGGCTTCCAAAGAGGGGCTGTGACCAAGCTCTCAC	221335
Qy	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr	800	Qy	1102	CTGTGATAGCTCAGAAATTGTGATGATCAGAAACTGGCCTGAGAAAGAAACCCAGAGA	221275
Db	222474	TACACCTGGCCATCAACCGCTCGGGAACAATGCAATGCCCAAGAGTCGACACACTAC	222415	Db	221394	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221215
Qy	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820	Qy	1102	GCCTGCCCCCTGTACAGACTGTGAGTGTGGGGGAAAGGCTTGTGAGGCTGCCCTGGC	221155
Db	222414	CTGGAAGATCTGTGAATATGACGTGAATGACAATGCCCCCAAGTTCCTCAGAGATTCCTAC	222355	Db	221274	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221095
Qy	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840	Qy	1102	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Db	222354	CAGGGAGGTCTCTATGAGATGTGCCACCTTCAACAGCCTCTCTGAGATCTACGCAAC	222295	Db	221394	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Qy	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly	860	Qy	1102	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Db	222294	GATCGGACACTCCGGTCTGAACCGGCAAGGTTTCTACACCTTCCAAAGAGGGGAGTGA	222235	Db	221394	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Qy	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgAlaLeuAsp	880	Qy	1102	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Db	222234	GATGTGACTTTATTTGATGATCAATCAAGGCAATGCTGTGCACTGCGGAGAGCTGGAT	222175	Db	221394	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Qy	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspTyrGlyMetProPro	900	Qy	1102	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035
Db	222174	CGTGAAGATGTGGCCCAAGTACATCTTGAGAGCGTACGCGGTGACAAAGGGGATGCGGCA	222115	Db	221094	AGACCTGTGGCTGACAGGGATGAGGTGTGAGGGGAGAGCAATCGTTCCTCAGACCTG	221035
				Qy	1102	TCCTGAGCTGGCGCAGGCCCCCGCTGTGCCCCAGGCTGGGCTGCTTTGGTTG	221035



Db	221034	TCACCTCATGGCAACCCTGTGATCTCA	CACAGCCTTTGGCAGGCTCAGAGGCTAC	CAGAG	220975
Qy	1102	-----	-----	-----	1102
Db	220974	ACACAGATTCACCAAGGGGGGTAGAG	ACCCTGCGCTCTCTCCCACTGTGTT		220915
Qy	1102	-----	-----	-----	1102
Db	220914	TCAGTAGCTGCAAGAACTGACTTTG	CTCTAGACATCTTTTCCCTCCAGGC		220855
Qy	1102	-----	-----	-----	1102
Db	220854	AGAAAGCCTGCTTGGCTCCCTTCA	GAGCTCCAGCAGAGAGTAGAGTCA	GCTC	220795
Qy	1102	-----	-----	-----	1102
Db	220794	ATTGCTCTGTCAGACAGACCTCA	CTGCCGAGTTCTTGGGTCTGGCTTG	GGG	220735
Qy	1102	-----	-----	-----	1102
Db	220734	GAGGGCTGCACAAAAGCAGGAGAT	GACTTTTGGGCTCCCCACCCCGTG	GT	220675
Qy	1102	-----	-----	-----	1102
Db	220674	TGTTCTTGGGGTCTAGGCTTCT	CTGCTGACATGATGGAGCTCAGAG	AAACAG	220615
Qy	1102	-----	-----	-----	1102
Db	220614	CACGGGTTTCAAGTGCATAGGACT	ACGCTTTGAGGGCAGAGACTCAG	GGCCCTG	220555
Qy	1102	-----	-----	-----	1102
Db	220554	ACTGTGTAGAGAGACAGATTTCT	AGGGGCTGTGGAGCCTTGCCAGAA	CCAGCTGG	220495
Qy	1102	-----	-----	-----	1102
Db	220494	GCATTCCTGCAGCTTGTTCAGAC	CTCAGCCCCCAGACAGAGACTGG	GGAGAG	220435
Qy	1102	-----	-----	-----	1102
Db	220434	CAGAGCTGGGTATGATTTCTAT	CCCTTCTGTACGCTGTGAGATG	AGCAAGCGG	220375
Qy	1102	-----	-----	-----	1102
Db	220374	GAGGGGAGACTAGTAGCAAGAC	CTTCTCTAGTCTGGAGTTGGGCT	GGAGAGG	220315
Qy	1102	-----	-----	-----	1102
Db	220314	AACGTTAAGTTTTCCAGGGGTG	AGACTATGTGCTTACAGGCA	GTCTTACATC	220255
Qy	1102	-----	-----	-----	1102
Db	220254	AGGGAGGGAAAGATGGCTTG	CGCCCTTCCGTCTACTTCTTAC	TGACACACTT	220195
Qy	1102	-----	-----	-----	1102
Db	220194	CCTGATCCACCCCTCAGGTGCT	GCAAGACAGAGAGAGCTCTCA	AGGCTATG	220135
Qy	1102	-----	-----	-----	1102
Db	220134	GCTTGTCAAGCTGTAAACCACT	TGAAGCTTCACTCCCTCCAGAG	AGAGAGCTAGGA	220075
Qy	1102	-----	-----	-----	1102
Db	220074	GACCCCTCAGCTGAGCTGGCA	CGAGAGCAAGAGCGCTCTCC	CGCTCTCCA	220015
Qy	1102	-----	-----	-----	1102
Db	220014	TCCACAGCTTCAGTGTGAAT	TCGGCTAGATTCCTAGGCGC	CGTGTGAGACAGCTCT	219955
Qy	1102	-----	-----	-----	1102
Db	219954	GCTTACCTTTGTCTTCTTCA	CAGGGGACACAGCCCACTGCA	CAACCTTAGATAG	218995
Qy	1102	-----	-----	-----	1102
Db	218994	GGAAAGCTGGCCGACGTAGGG	CACTGTGCTCTTCTGCGCTT	GTGCCCCCTCC	218935
Qy	1102	-----	-----	-----	1102
Db	219834	CCATCAGCCCTGACCTTAGAG	TCAGTCAACCAAGCAACCCAA	AGTTCAGAGGCTGG	219775
Qy	1102	-----	-----	-----	1102
Db	219774	TGAGAGGACTGACGCCCAT	GTCTCTTGTCTCTGTGTTGT	CTCTCT	219715
Qy	1102	-----	-----	-----	1102
Db	219714	TCCCTCCACCTGCGCAATGG	AGACCTCTGCACTGCTGTCT	GTACCTTCCCCAG	219655
Qy	1102	-----	-----	-----	1102
Db	219654	CTTCTGATGCTTTCCTTCT	ATAGCTTCTCACTGTGCT	TTGTCTTCCGTTC	219595
Qy	1102	-----	-----	-----	1102
Db	219594	TCTCTATGTTAGATCTGT	CTTCTCTCTCACTGTGCT	GAGCTGCCAGCTGG	219535
Qy	1102	-----	-----	-----	1102
Db	219534	CTCCATGGCTCTGACGCC	CAATGTGTCGGCTGTATCT	CTTCAACACAGAGAA	219475
Qy	1102	-----	-----	-----	1102
Db	219474	CTTGCACAGGCTGTCACT	CTCCAGACAGAGTCTTAT	CAAGAGAGAGTTG	219415
Qy	1102	-----	-----	-----	1102
Db	219414	GAGGCTGACTGTGGCT	CTTCTCTGTGAGCTGT	AGGCTTAGACCAAGTAT	219355
Qy	1102	-----	-----	-----	1102
Db	219354	TGCTGGGTGAGGCTTAC	CCCAAGCTCTCTTCTTAA	ACAAGTCAGGCCACTCTATTC	219295
Qy	1102	-----	-----	-----	1102
Db	219294	CCAACCTGCTTGGAAAG	AGACAATAGAGAGAGAG	AGCAATGGTTGCTG	219235
Qy	1102	-----	-----	-----	1102
Db	219234	GATGGGAGATGACAGAG	GGGCAAGTGCACAGAT	GTGGCCAGATGGGAGAGAC	219175
Qy	1102	-----	-----	-----	1102
Db	219174	AAAGATGCCCCGGCAAG	GGGACAGTCAGGCTGTGG	CTGTGAACCATGGGGAGTG	219115
Qy	1102	-----	-----	-----	1102
Db	219114	TGGCTTCCAGGACAGAG	ATGCAAAATCTTGTCTGT	GTCTCAGAGACTGT	219055
Qy	1102	-----	-----	-----	1102
Db	219054	TTGTAGTCCCTGTGTT	CTTCTGTGAGAGGAA	CATCCATCCAGAAACATGGGC	218995
Qy	1102	-----	-----	-----	1102
Db	218994	TGTATCTTCACTCTGAG	CCAAATGCAAGGCTT	TGACAGAGGCAAGAGTCTGATC	218935
Qy	1102	-----	-----	-----	1102
Db	218934	ACCATGTGACCATGAG	CGAGCTTGAGGGCTT	AGAGGCTTGAATGTCTGTGTA	218875
Qy	1102	-----	-----	-----	1102
Db	218874	CCCTGGCAAGACAGAG	AGAGGAGTAGACGAC	GTCAATGACAGGCTTGTAGAA	218815



D	21654	GCTCACCCGCANTCTGAGCTCAGCGGTCTCCCTTCGACACAACTTTGCTCGCAGAA	216595
Q	1235	UProCyseUantYrthelchagCyseValSerValleuArpheAseSerAlaProph	1255
D	21654	ACCCGCGAGAAATTAACATGCGGTGTGTCCGTGCTGCGCTTCCACTCTCTGCGCCTT	216535
Q	1255	e1leA1seSerSerValleuPheArpPro1leH1ProVal1leU1leuArpCybar	1275
D	21655	TATCCCTTCTCTTCTGCTGCTCTTCCGCCCATCCATCCATCCGCGGGCTGCGCTCTCG	216475
Q	1275	gCyseProProglYpethrgrlYasPrYCyseglUthrglUValAspleuCybrYseAr	1295
D	21647	CTGCCCAACAGGCTTCAACAGCGACTACTGCGAGCTGAGGTGAGCCCTGTTACTCAAG	216415
Q	1295	gProCyseglYProH1seglYArYgCybarSerArpGlUglYglYrthYrthCyseUeCybar	1315
D	21641	ACCTGTGTAGACCCCAATGGGCACTGCGCGAGTCGAGAGGGGTGTATACCTGCTCTGTGCG	216355
Q	1315	gAsp1YrYrthr-----	1319
D	21635	CGATGGCTACACGGGTGTAGTCTAGGGGCAAGGACATGGGGGCAAGCCCTGCAGGGCTGACG	216295
Q	1319	-----	1319
D	21629	CTGTGGCCAAAGCAAAATCAGACTAATATATGGGCTGCTCATTACTTCCAGAGGTATA	216235
Q	1319	-----	1319
D	21623	AACTAGATGCAACAGCCGCTTTAGATCTTACTGTGTTCACACTGCTGCTTAGTA	216175
Q	1319	-----	1319
D	21617	AGTCAGATTAAGAACTTATGTCTACTGTGCAAGCCCAACACCACTGGGCAATGC	216115
Q	1319	-----	1319
D	21611	CGCCTGGCAACGCCCAAGAACTGTGTAATTAGAGTCAGTCCGCGACCTGGGTTTGT	216055
Q	1319	-----	1319
D	21605	GTTCCAGCTTACCACTAATACTGTGAATAATGAGCAACTACTTACTTCTTCCAC	215995
Q	1319	-----	1319
D	21599	AGTATTTAAGATGAGATAACCAATTATGTAGGTCAAAAGTTGCTGTGAGACTCATG	215935
Q	1319	-----	1319
D	21593	TAAATTAATTGTATGAGAAAGAAAGCTGCTTATGTGTGAGTGTGGCTTTGTTTGAAG	215875
Q	1319	-----	1319
D	21587	GAAGTCTGAGATGAGTGTGATTCAGTTAAGAGTGTGTGCTGGCTTAAGAACAGAGG	215815
Q	1319	-----	1319
D	21581	CTTCATTAATGCTGTGTGAATCCTTCAATTACTTTAGCCAGACTTCCCAACCAAGAA	215755
Q	1319	-----	1319
D	21575	CTGAGAGAGGCACTGAGAAAGCAGGGGTTTGAAGTCAATGGGAATTCAGAGCACTGAGAC	215695
Q	1319	-----	1319
D	21569	CGTGTAGCCCTGCAAAATGCTCTCATTTCTAATCCCAATTGTCTGTGGGCTGCAAGTCT	215635
Q	1319	-----	1319
D	21563	CGTTACACCTGCCCAACCTGTAGGCTGTGCTTCTCTCAGCTCTTGTCTGTGTTCC	215575
Q	1319	-----	1319
D	21557	TTCCAGACTCCCAAGCTTCAAGTCTCTCAGGCCCAAGCGCACTGGCTCTGCCAGTGGCG	215515
Q	1319	-----	1319
D	21551	TGGGCAAGCTAGTGCCTCTGTGACACCTCTGTCTCAGAACTTGGCAAGCTTGACTCT	215455
Q	1319	-----	1319
D	21545	TCTGTGCTGTGTGTGTGAGGTGTGGGTGTGGGGCTAGGGAGGGGGGAGTCTTGTGCC	215395
Q	1319	-----	1319
D	21539	AGGAGATGGGGAAAGGGGGCTCCGAAAGTATAGCGGAATTAACATTTCCATGTAAACAAGA	215335
Q	1319	-----	1319
D	21533	GCCAGTCTAGGCAACAGGAAGTGAATCTATCTTAGAATCAGATCGGAGGAGGA	215275
Q	1319	-----	1319
D	21527	AGGAGTGATTTTCCGCTCTAATATCTGAATTTTCAGCCCCCTGATTTCTGTGTCC	215215
Q	1319	-----	1319
D	21521	TCAGGCTTAGGCCCCCAAGTCTCTCAGCCTGTTTAGCGCGGAGGAGTCTCAGTTGAGA	215155
Q	1319	-----	1319
D	21515	CCCAATGTCTGTTCTTTGGGAAGCCCTGCTTACTTCAACACATGACCCAAACTAT	215095
Q	1319	-----	1319
D	21509	AACTCAATTAGACTGGAAACTATGTCATCTCTAGGAAAGAAACTGATCTCTGACC	215035
Q	1319	-----	1319
D	21503	TGACAGAAACAAAACAGAGGTGGAAAGGGGGGAAGAACGGAGTGGGCGCGCTG	214975
Q	1319	-----	1319
D	21497	GAACTGTGAGGGGCAAGGCTTAGGGAACAGGGTACTGACAGGTGTGGGAAAGGGCCC	214915
Q	1319	-----	1319
D	21491	AAGACTCACAGCAGAACTTTGTGAGCTGACAGAGAAATAGTGTGCGAGCCAAATTAAG	214855
Q	1319	-----	1319
D	21485	AGGAGTGTGTGCCATAAATTTACAGAGATGTCAACAAAGAAATGAGAACGATGGGA	214795
Q	1319	-----	1319
D	21479	GCCCCAGAGATGAGAGAGGTCAACAGAGATTTTACAGGGACACGTTAGACTTCTCACT	214735
Q	1319	-----	1319
D	21473	ACTGAAATTAACCAAGAGAGACAGTTTAAACAAAGAAAGTTTACTTGTCTCCGCTCTC	214675
Q	1319	-----	1319
D	21467	TGCAGTTTCAGTTCTGTGTGCTGTGTCTTCTGCTATTAAGACAGACATCAAGGTGTGGAAAG	214615
Q	1319	-----	1319
D	21461	CTGTCAACCAAGATGAGAGAGAACCTTGGGACAAAGATGTAGCCTGTATGGCTACTCCA	214555
Q	1319	-----	1319
D	21455	GTAAGTGTCTTCTCAGCTCTCTCTCCGATGTCTTCCAGTGTCTCATTAACAGCCCC	214495
Q	1319	-----	1319
D	21449	TCAGCAGACCAAGCCTTTAGATATAGAGCTGGGGAGGGGGCAATTCATATTCGAAC	214435

QY 1319 ----- 1319  
Db 214434 TGTAAAGGAGACCTTCCAGCATCATCTGACAGGCTCACTGCTTGCCTTCCCTCC 214375  
QY 1320 -GlyValHisCysGluValSerAlaArgSerGlyIleProGlyValCysAlaHis 1339  
Db 214374 AGCGAGACCTGTGAAGTGAATGCTCCGCTCAGAGCCGCTGTGATCTCCAGAGTCTGCAAGA 214315  
QY 1339 nglValThrCysValAsnLeuValGlyIlePheIleCysAspCysProSerGlyHis 1359  
Db 214314 TGGGGGACCTGTGTCACTGTGTGGAGGCTTCAAAATGACTGCTCCATCCGAGGA 214255  
QY 1359 pPheGlyProIleCysGluValThrIleArgSerPheProAlaHisSerPheIleTh 1379  
Db 214254 CTTTGAAGACCTTCTGCAAGGTGACCAACGCAAGTTCCAGCCGCTCTTCAATCAC 214195  
QY 1379 rPheValGlyIleValArgIleValPheHisPheThrLeuAlaLeuSer----- 1394  
Db 214194 CTTCCGGGCTTCCGCAAGGCTTCCACTTCCCTGCTCGAGTGAAGAAGCC 214135  
QY 1394 ----- 1394  
Db 214134 CTGGGCTCCAGGACCTTGTGAAGCGTGGGCTCTCTGTCCAATGAACGGGGTGG 214075  
QY 1394 ----- 1394  
Db 214074 CAGATCAGTAAAGAAAGTACAGAAAGCAGTGGCTGACAGAGCACTTGTATGCTTGGCA 214015  
QY 1394 ----- 1394  
Db 214014 GCGCGCGCGCTCACTCTCAGAGTCTGGCTGACCTCTCCCATCTTGAAGATGGGCG 213955  
QY 1394 ----- 1394  
Db 213954 GCCAAGATGCTTGGGCTTCTGATCTGACTCAGAGCTTAAAGCAAGTCACTGTGC 213895  
QY 1394 ----- 1394  
Db 213894 TGTGTAGATGGCCAGGATGACACAGGGGTGGGTCGCCGTTCTCGGCCCGTT 213835  
QY 1395 ----- PheAlaThrIleGlyIleValArgGlyIleLeuLeuIleValArgGly 1408  
Db 213834 GACCATCCCTTATACAGATTTGTCTACCAAGAGCGGATGCTGTGTACACGG 213775  
QY 1408 ValArgPheAsnGlyIleValHisPheValAlaLeuGluValIleGlnGluValGlnLe 1428  
Db 213774 GCGCTTCAATGAAGATGATCTGTAGCTCTCGAGGTATCCAGAGGAGGTGAGCT 213715  
QY 1428 uThrPheSerAla----- 1432  
Db 213714 CACTTCTCTGCAAGTGAAGCCAGCGCTCTGTCTTGGCCATCTCTAAGATCTTATG 213655  
QY 1432 ----- 1432  
Db 213654 TCTCCAACTCTGACCTCAAGCAATGCGCTCCCTCCAGTCAAGCTTGAACCTGGCT 213595  
QY 1432 ----- 1432  
Db 213594 CCACTCTTAAGATGCCACTTCCCTGCACTCCCACTCTTCTCTGTACCCAGCTTAC 213535  
QY 1433 ----- GlyIleSerThrThrValIleSerProPheValProGlyValHis 1448  
Db 213534 CGGCTCTTTTCTTGGGGAATCCACCAACAGGTGTCTCAATCTGTGCTGAGAGGGTCA 213475  
QY 1448 eValArgGlyIleThrIleThrValIleLeuIleValIleValIleValIleValIle 1462  
Db 213474 GTATGGCCAGGTGCAATACGTACAGCTGAAGTATCAATTAAGTGGATGATGGGGGGA 213415  
QY 1462 ----- 1462  
Db 213414 CAGAGAGCTAAGAGGCTCTGTCCGTGGCTGTATTTGTGATGTGGGGAAGGCTGTTC 213355

QY 1462 ----- 1462  
Db 213354 CAGCGCTGGAAAGAGGAAGATGTGCTTGCAGAGGCACTTCCATCCCTGTGTGCT 213295  
QY 1462 ----- 1462  
Db 213294 TGATGTGGAAATATGTGAAGCGAGGTGCTCTGTGGTGTCTACTGCTTCTT 213235  
QY 1463 ----- ProLeuLeuGlyIleThrIleValProGlnGlyProSerGlyIleHis 1478  
Db 213234 CTGTGTCAAGCCGCTGTGGGTCAAGAGGACTTCCACAGGCCATCTGAAGCAGAG 213175  
QY 1479 ValAlaValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerVal 1498  
Db 213174 GTGGCTGTGTGTCTCCGTGAATGGCTGTACACAGGGGTGGCTGTGCTGCTGAGACTATG 213115  
QY 1499 LeuGlyAsnIleYrSerCysAlaAlaGlnGlyIleThrGlnGlyIleSerIleVal 1515  
Db 213114 CTGGGCAACTACTCTCTGTGTGCTCCAGGGCAACCCAGAGAGCGACAGAAATGAACAGAG 213055  
QY 1515 ----- 1515  
Db 213054 ATGGGCGGGGAGGGGAGGGGTGACAGAGCTAAGGAGAGGTATGTGGGTCAACAG 212995  
QY 1515 ----- 1515  
Db 212994 ATGGGGCTCCAGGTCAACAGATGGGGCTCCGAGTGAAGAACTACTAGATGCAACCTT 212935  
QY 1515 ----- 1515  
Db 212934 CCTTCAATTTGGGAATTAAGTCAAGTCTTGTGGCTGTGAAGGAAAGGAGCAATC 212875  
QY 1515 ----- 1515  
Db 212874 TGTCAGTACTGGGGTGGCTGTGCTTCTCGGGTTTCTTAAAGGCTCATGCGCTCG 212815  
QY 1515 ----- 1515  
Db 212814 GCTCAAGTCTTCTTCTGTGCTTGAATGCAAGGTATGGGTAAAGAAACCAACAG 212755  
QY 1516 ----- SerLeuAspLeuTh 1520  
Db 212754 ACTTAAATCTGTGCTGCTTCTTCAACTCTTGTCCCAAGTCTCTGACCTGAC 212695  
QY 1520 rGlyProLeuLeuGlyIleValProAspLeuProGlySerPheProValArgMetAr 1540  
Db 212694 AGGGCTCTGTGTGCTGGGTGGGTGCTGAGATGCTGCCAGATTTCCCTGTCCGAATGG 212635  
QY 1540 gGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAs 1560  
Db 212634 GCACTTGTGGGCTCATGAAGAACTCCAGGTGATAGCCGCAAGTCCAGATGATGGCCGA 212575  
QY 1560 pPheIleAlaAsnArgIleThrValProGly----- 1570  
Db 212574 CTTCAATGCAACAAATGACATGTGCTGG-TATAGAACCTGGGGATACGGGCAAGCT 212516  
QY 1570 ----- 1570  
Db 212515 GGGAGCAAAATGCAATGAAGATCGGTGTGATGACAGGGAAATGCAAGCTGCGGTGT 212456  
QY 1570 ----- 1570  
Db 212455 CTTCCGCTCACTGTGTGTTTCCAGAGCCCTGGGGTCCCTGAATCCCTCTCTCCGT 212396  
QY 1571 ----- CysProAlaValIleValAsnValCysAspSerAsnThrCysHisAsnGlyValTh 1588  
Db 212395 TCTAGGCTGCCCCACCAAGAAACGTGTGTACAGCAACATTCACATTAATGTGGCAC 212336  
QY 1588 rCysValAsnGlnIlePheAlaPheSerCysGluCysProLeuGlyIleGlyIleValHis 1608  
Db 212335 CTGTGAAACAGATGGAGCGCACTTCAAGCTGAGAGTGTCCCTAGGCTTCCGGGGCAAGAG 212276  
QY 1608 rCysAlaGln----- 1611

```

Db 212275 CTGTGCCCCAGGGTAGAGATGGACGCTGCCACAGGTTGGGGCCTGGAGCTCTGTCACTTGTTC 212216
QY 1611 ----- 1611
Db 212215 TGGGATTACGGGAGAGGCTTGGGGCAGGCTCTGGGCGAGGGCTCCGGAGTCAAGGCTGTGG 212156
QY 1611 ----- 1611
Db 212155 GTGGAGAGCGGTGTCTGGGCAAGTAAAGAGAGCTAGGCAAGAGCAAAAGACCTAGTC 212096
QY 1611 ----- 1611
Db 212095 AAGAGCAGGCGCTGGGCAATGTAGGGCGGCGCTGGAAAGAGTAAGTCTCCGGTGGGC 212036
QY 1611 ----- 1611
Db 212035 GGGGCAATAGCAGAGACCGCGGTGTAGAGGGAGGAGGTCATCTCTCATGGAGCTGCCA 211976
QY 1612 ----- 1614
Db 211975 CGCTTCTGATCCTCACTGCCCTGCTCCCTTGAAATCACTTTTCATCTGCTCCAGAAATGG 211916
QY 1614 1AAsnProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProI 1634
Db 211915 CCATATCCCAAGCGCTTCTTGTGGGAGAGCGCTGTGGCCCTGGGCAATGACCTCTCTGCCCCA 211856
QY 1634 1eSerGlnProTrpTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuL 1654
Db 211855 TCTCTCAGGCGCTGGGACCGTCAAGCTCATATGTTCCGGACAGCGCAAGAGATGGCGTCTGTC 211796
QY 1654 euGlnAlaAlaIleThrArgGlyValArgSerThrIleThrLeu----- 1666
Db 211795 TGCAGGCGGTCAACAGAGGGGCGCAGGACCATCACTCGAGGTGATGACAGGCGCGTGGG 211736
QY 1666 ----- 1666
Db 211735 AGTGTGGCAAGCCCGGCGTGGCTAACTTTGTAGAGACCTTGAGGGGCTTGGACCCAGGT 211676
QY 1666 ----- 1666
Db 211675 GGCTGGGCGAGGTTTGGCAGAGCCAGACATTTGGGTGGCCCATTCGGGTGGTCACTACT 211616
QY 1666 ----- 1666
Db 211615 GCCACCTATTGGGCTTAAAGAAATATATCCCTCTGCTGCCGAGCGCGCAGACTGG 211556
QY 1667 ----- 1671
Db 211555 AAGAGCCTACTGAGGTGCTTAATAATTTTTTTCTTCCCTCTCTGTAACGCTTCGGGCGAG 211496
QY 1671 1HisValMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnP 1691
Db 211495 GCCACGTAGTGTCTAATGTGTGAGGGGACAGGGGCTCAGGGCCCTCATCCCTGGGTCTGGAGC 211436
QY 1691 roGlyArgAlaAsnAspGlyAspTrpHisHisAlaGlnLeuAlaLeuGlyValAspGlyG 1711
Db 211435 CAGGCGGAGCCATATATGTGTGACTGGCAGCAGCTGTGCACGTGGAGAGGTACCGGGG 211376
QY 1711 1PProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuG 1731
Db 211375 GCCCTGGCCACCGCCATCTCTGCTTGAATGTGGGCAACAGAGGAGGAGTAATCTGG 211316
QY 1731 1PProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyGlyLeuProGlyProA 1751
Db 211315 GCCCTCGGCTGACCGGGCTGCACCTGAGCAATATAGCAAGTCCGGGAGAGTCCCGGCGAG 211256
QY 1751 1aGlyGlyValAlaArgGlyPheArgGlyCys----- 1761
Db 211255 CCAGCAGTGTGGCCCGGTGGCTTCGGGGCTGTTTGACAGGTGAATCTCTCCCGCAGACTT 211196
QY 1762 ----- 1763

```

```

Db 211195 CCATCCATCTTCACTGGAGCCCTGTTCCACTGAGGGCTAATGCTCTTAATCCCTTC 211136
QY 1763 1nglyValArgValSerAspThrProGlnGlyValAlaSerSerLeuAspProSerHisGlyG 1783
Db 211135 AAGGTGTAGGGTAAAGCAGACACCCAGAGGTGTAGAGAGTGTGATCCAGCCGTGGGG 211076
QY 1783 1uSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsnProCysP 1803
Db 211075 AAGACATCAATGTGAGGACGAGGCTGTAGCTGGCCAGATCCCTGTGACTCGAATCATATCTC 211016
QY 1803 roAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp----- 1820
Db 211015 CCACCAAGCTACTGACAGCAACACTGGAGCAAGCTAATTTCTTGTAGCTGTATCCAGGTA 210956
QY 1820 ----- 1820
Db 210955 AGCTGAGAGTCTGGAAATAGGGGCTGTGTGCACGGGTCACATCCGCTCTAATATGTA 210896
QY 1820 ----- 1820
Db 210895 GCCAGTTACATCCAGGTAATAGCCATCTGTGTGACAGGTTCCAGACGGAAGTAATAAGTA 210836
QY 1820 ----- 1820
Db 210835 GTTGTAGAGTGAAGAAAGATATGCCAGTAAGGGCGGGTCTTGAACTTTCCGATCTTCTT 210776
QY 1821 --ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisG 1840
Db 210775 GTCCAGGTTACTAATGTGTGACACATGTATATGTGTGTGACTGTAAACCATGCGAGCACCC 210716
QY 1840 1nSerValCysThrArgGlyAspProSerAlaProHisGlyTyrThrCysGlnCysProProA 1860
Db 210715 AGCTGTGTGTATCCCGAAACCCAGTGCACCCGAGCTACATCTGAGAGTGTATACAA 210656
QY 1860 snTyrLeuGlyProTyrCysGlnThrArgIle----- 1870
Db 210655 ATTACTTGGGCCCTAATGTGTAGACACAGTA--GGTGGCGGGGGCTTAACCTCTGTATCAC 210597
QY 1870 ----- 1870
Db 210596 TCAGCCCCGAGAGTCAGAGATCGGGCTGAGCTGAGCCAAATGAAGAATAATGTGGTT 210537
QY 1870 ----- 1870
Db 210536 GAACCAAGCGGAGAGATGCAGTGTAGAGGGGCTCATGCTGTGGTGTCTTCT 210477
QY 1871 -----AspGlnProCysProArgGlyTyrTrpGlnHisProThrCysGlyProCysA 1888
Db 210476 GCAGGATTAACCAACCTTGCCCCCGTGGCTGTGTGGGAGACACCCCAATGTGTCCATGCA 210417
QY 1888 snCysAspValSerTyrGlyPheAspProAspCysAsnLysThrSerGlyGlyCysHisG 1908
Db 210416 ACTCGAATGTACAGAAAGGCTTTGACCACAGACTGCACAAAGACAAAGAGGGGAGTGCACCT 210357
QY 1908 ybLys----- 1909
Db 210356 GCAAGGTGAGAGACCCGAATGAGCCCCCGTGTGTATCCCTGTCTGTAGCTTCTGCCCC 210297
QY 1909 ----- 1909
Db 210296 TACAGGGGAAAGAGAGATGGGCTTCTCTCAAGAGCGAGAGCAATCCTGTGCTT 210237
QY 1909 ----- 1909
Db 210236 TTGCTTATACAGGATGAGAGGTGTGTTTGCCCTTAATATCCGCTGCAGGGGAGATCCA 210177
QY 1909 ----- 1909
Db 210176 GTTCCAGTGTAGTACGTAGAGCTGTGACTAATGTGTGGCTGTTTTCAGAGCCATTTCTG 210117
QY 1909 ----- 1909
Db 210116 AGTATTTGGCTTCTCACTGCTATGTCAATTACCTGCTGTGTAGAGTGTGAATCTCTCCC 210057

```

QY	1909	-----	1909
Db	21056	CACAAGAACCCAGCATAGGGAGGCTCTCTGAGGAGGCCACAGCTTGAGCTTGCT	209997
QY	1910	-----	1913
Db	209996	AGCATCCAGAGAAAGACATCTTGATCCAGTACCCCTTTCTTCCCTCCTAGAGAAATCAGT	209937
QY	1913	YTAATPProglYserProthCybLeuLeuCybAspCybYrProthGlySerLeuS	1933
Db	209936	ACCGACCCCCACAGACGCCACTTGCTCTGTGTGACGTACCCACACTGCGCTCTTGT	209877
QY	1933	erArgValCybAspProglYAspGlyGlnCybProCybYsProglYValIleGlyArgS	1953
Db	209876	CCCGAGCTGTGACCTGAGAGATGCGCAAGTGCATGCAAGCCTGGAGTCAATTGGGCGT	209817
QY	1953	IncYAspAspArgCybAspAspProPhaIleValIleThrIleAnGlyCybGlu-----	1970
Db	209816	AGTGTATGCTGTGACAAACCTTTGTGAGGTCAACCAATGGCTGTGAAGGTAGG	209757
QY	1970	-----	1970
Db	209756	GGGCTTTAGACAGGTAGTCCCTTCCTCCATATGCCAAGAGCTTACACTCCAGAG	209697
QY	1970	-----	1970
Db	209696	TTAGAAATGAGCTGGCATTTGGGTGGGAGATGGAAGAGAGGTGCGGGGCATGGATAG	209637
QY	1971	-----	1979
Db	209636	AACTGCTTACCGATTAACCATGTGTTCTTCCATATGATTAATGACAGTGGCCACGG	209577
QY	1979	IaIleGlnIleGlyIleTPTPTProArgThrArgPheGlyLeuProIaIaIaIaProc	1999
Db	209576	CCATTAAGGTGGGATCTGGTGGCCCCGACGGGTTGGGCTTACCTGCTGCGCCCCCT	209517
QY	1999	YbProYleGlySerPhe-----	2004
Db	209516	GGCCCAAGGGCTCCTTGGTGTAGTTTATGATTAACAATATGATATGACAATTGACCTT	209457
QY	2004	-----	2004
Db	209456	ACTCTAGAGAAATAAGATAGAGAGTCCAGCCAGCTTGAAAGCTCCCGGGCCCACTT	209397
QY	2004	-----	2004
Db	209396	TCGTGATTTCTATCTCCAGAGACTCAGTGTCTCACTTCACTTCACTCACTCACTCAA	209337
QY	2004	-----	2004
Db	209336	CCTAGGCTGACCAAGCTGCTTATTCAGCCACACAGAGTGAAGAAATCTGCGTG	209277
QY	2004	-----	2004
Db	209276	TCCTGAACAGGGCTGGAGCTAGACTGTGTGATGTGATGGGGTGCTTATCTTGAGCTAAC	209217
QY	2004	-----	2004
Db	209216	TACTTAGTACTGAGTTAACCCCACTGAAGAGGCGGTGCTGAGAGATATGGCGTGCT	209157
QY	2004	-----	2004
Db	209156	GGGTACTATCCAAAGATTCAATTAATGTGTGACATGTTTCATGGGCAAGGAGTAGCT	209097
QY	2004	-----	2004
Db	209096	GAGAGTAAGGCACTAGCTGATGAGACAGGCGTATGATATCTTGGGGCGGTGTGGCACA	209037
QY	2004	-----	2004
Db	209036	GGTTACATAGCGTCCCTTTGGCACTGTACTGATTTTGTGAGATGTTTTGGCTGTGAG	208977

---

QY	2004	-----	2004
Db	208976	CTCTGTACACTGGGCTTCTCTGTCTCTCCCTAAGCCCCCGCTTCTTGCTTCTTCTCA	208917
QY	2005	GlyThrAlaValArgHisCybAspGlnHisArgGlyTPTLeuProProAsnLeuPheAsn	2084
Db	208916	GGGACTGCTGTGCGCCACTGTGATGAGCACAGGGGCTGCTCCGCCAAACCTCTTCAAC	208857
QY	2025	CythrSerIleThrPheSerGlnLeuYsGlyPheAla-----	2037
Db	208856	TGCAGTCAATCACTTCTCAGAGCTAAAGGCTTGT-AGGTGACCCCTTGTCTTCTC	208798
QY	2037	-----	2037
Db	208797	CTCTTTCATACCTCGTCTTTGAAAAATAGAGTCCCGGGGTGGGATTTAGCTCAGT	208738
QY	2037	-----	2037
Db	208737	GGTAGCGGCTTGCTTAGAAGCGCAGAGCCCTGGGTTCGGTCCAGCTCCGGGAAAAA	208678
QY	2037	-----	2037
Db	208677	AGAACCAAAAAAAAAAAAAAAAAAATAGAGCCGAGATCTGTCTTGAGGCTCT	208618
QY	2037	-----	2037
Db	208617	CCTACCTGAGAGGCCACTTCTTCCCACTGATCTTTGAGGCCCTTTTACTGTAGCC	208558
QY	2038	-----	2039
Db	208557	CCAGTCCCAACATCTGTGTGACCTCTGCACTTGTGTCTTCATCCAGGCCAGGG	208498
QY	2040	LeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnLeuAlaLeuLeu	2059
Db	208497	CTACGCCAAAGCATCAGGCTGAGCTCAGGACCTCCAGAGGCTGGCTGCTCTG	208438
QY	2060	ArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIysValIaTyrGln	2079
Db	208437	CGTAATGCAACACAGCACACCTCTGGCTATCTGGCAATGATGTCAAGGTGGCTTACAG	208378
QY	2080	LeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThr	2099
Db	208377	CTGGCCACACGACTGTGCTCATAGAGTCCAGCGGGCTTGGGCTGTGCCACACA	208318
QY	2100	GlnAspValHisPheThrGlu-----	2106
Db	208317	CAGAGTGTGCACTTCACTGAGGTGGGCTTGAAGTGAAGCATGTGGGCCAGTAGAGA	208258
QY	2106	-----	2106
Db	208257	GCCTGTCTCGAGATCTTGGGGAGATCTTGAAGGAGATCTTCTTACTGTAGAGAC	208198
QY	2106	-----	2106
Db	208197	TGCCCAACCCCTGGTACAGAGGCTTCATGATATGTCTGTAACCAAGCGCACACAGTTG	208138
QY	2106	-----	2106
Db	208137	ATTAGAGTACCTCTCTGGGGCTGGGAAATTAGCTCAGTGTAGACGCTTACTAG	208078
QY	2106	-----	2106
Db	208077	AAGCGAAGCCCTGGGTTGGTCCCGACCTCCGAAAAAGAAAAAGAGTA	208018
QY	2106	-----	2106
Db	208017	CCTCTCTTACAGTGTCAAGATTAATTAAGTACCTAGCTTGAAGGCTGACCTGAGCA	207958
QY	2106	-----	2106
Db	207957	GAGCCCACTATTATACCAAGAGCTCAGTCTTGTCAAGGAGCCAAAGGACAAACA	207898
QY	2106	-----	2106



```
Dh 207897 GAGCGAGACATGCAGAGACACATGGGTGGCCCTGATTAAAGTGTGAAGGAAGTTTAA 207838
Qy 2106 ----- 2106
Db 207837 GGGCGAGTACCATGCTGGAAGTGAACAGTATGATACAGAGCTTCCAGGGGAACTG 207778
Qy 2106 ----- 2106
Db 207777 TGGGAGCTGTAAAGGAGAGACCTAGTGGCGGGTCTGAACAGGACAGAGCTGAACCTT 207718
Qy 2107 ----- 2107
Db 207717 GGTATAGGTTGACCAAGACATATGCCACTGTGACCAAGCCCATCTCCAGAAATCTGCTG 207658
Qy 2110 ArgValAlGlySerAlaLeuLeuAerThrAlaSerValArgHisThrGluLeuLeuLeuGln 2129
Db 207657 AGGTGGGAGAGCGCCCTCCCTGGATGGCGCCAAATBAAGGCACTGGAACTGATCCAGCAG 207598
Qy 2130 ThrGluGlyGlyThrAlaTrpLeuLeuGlnHisThrGluValAlaTyrAlaSerAlaLeuAla 2149
Db 207597 ACAGAGGGTGGCACTGCTGCTGGCTGCAGCACTATGAGGCTTAATGCCAGTGGCCCTGGCC 207538
Qy 2150 GlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThrProAsnIle----- 2167
Db 207537 CAGAAATATCGGCAATACCTAAGAGCCCTTCACTATGTCACACCCAAACATTGGTGA 207478
Qy 2167 ----- 2167
Db 207477 GTGGTCTGGGGCTGGGGGGCGGGGTGTGAGAGTCTGCCCATTTCAAAGGCGAGCCAG 207418
Qy 2168 ----- 2168
Db 207417 CCGTGTCTGGTGAAGTCACTTGCACCAATATTTCTTCTCTAGTACTCTGTGGTGGCGCTG 207358
Qy 2175 AspLysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGluGln 2194
Db 207357 GATTAAGGGAACTTGTCTGAGACCAAGCTGCCCGTTATGAGCGCTTAAGTGGGGAGCCG 207298
Qy 2195 ProProAspLeuGluThrThrValIleLeuProGluSerValPheArg----- 2210
Db 207297 CCTCCAGATCTTGAGACCAAGTCACTTTTGGCCAGAGTCTGTCTTCAGAGTCAAGCGGGGA 207238
Qy 2210 ----- 2210
Db 207237 ACGTGAATGTGTAACGCTGGGGGTGGGGCAACCCATGCAGTGGCGAGGCGCAGTGAAT 207178
Qy 2211 ----- 2211
Db 207177 GTAGCAGCATTTAATTGCAAGATGCTCCCAAGTGAATCTGCAGGACCTGGTGAAGC 207118
Qy 2224 agLngluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGly 2244
Db 207117 CCAAGGAGACTGAGAGAGCTGGCCCGGGGCGAGCGGAGGCAACCAAGAACTGAAGTCAAGGCGA 207058
Qy 2244 uAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAs 2264
Db 207057 GGCAGGTGGCAGATGTCACTATTACCAACAGCTGGCTGAGTGTGCCCAACATACGA 206998
Qy 2264 ProAspLysArgSerLeu----- 2270
Db 206997 CCGAGTAAGCGGTAGCTGAGGTAAAGCGGCTACGACAGGCTTAAGAGGTTGGGGTTAGC 206938
Qy 2270 ----- 2270
Db 206937 AAGGAATGTGAAGCATCTCCAGTGAATAGGGTGTCTCGACGCCGTTTGTGACTGT 206878
Qy 2271 ----- 2271
Db 206877 GTACCTTCTGTCAATCAGAGTCCCAAGCGCCAGTCAATCAACACCTGTGTGAAGCAT 206818
Qy 2285 eSerValHisAspArgGluGluLeuLeuProArgAlaLeuAspLysProValThrValG 2305
|||||
Db 206817 CAGTGTCCAGATGATGAGAGCTCTCTGCACGCGGCACTGGAACAAGCCAGTCAAGTGA 206758
Qy 2305 nPheArgLeuLeuGluThrGluGluArgThrLysProIleCysValPheTrpAsnHisSe 2325
Db 206757 GTTCCGACTGTGAGAGAGGAGAGAGCAACAGCCATCTGTGTCTTGTGAACCATTC 206698
Qy 2325 rIleLeu----- 2327
Db 206697 GATCCGTGAGGCTGCTGTATCAACCCCTTAAGCCATGACTTTGAGCAAGATCCAGGC 206638
Qy 2328 ----- 2328
Db 206637 TCCGTGGCTGTCACTTTCTCTCCCTGACTCCGTCTTCTCTGCAAGGCTCAAGTGGCA 206578
Qy 2331 hrcGlyGlyTrpSerAlaArgGlyCysGluValValPheArgAsnGluSerHisValSerC 2351
Db 206577 CAGGTGGCTGGTCCCGCAAGGCTGAGAGTGTGTCTCCGTAAACAGAGCCAGTCAAGCT 206518
Qy 2351 ySGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArgArgGlu--- 2369
Db 206517 GCCAGTGCACCAATATGACAAAGCTTTGCTGTGTCATGAGTGTATCCGAGAGAGGTTTG 206458
Qy 2369 ----- 2369
Db 206457 GACTCCAGGGGTAGCCCAAGAGCAGCGGTGGGCAAGTGCAGAGGCTTCAGGCCCTG 206398
Qy 2370 ----- 2370
Db 206397 CTCTTTTGAACCCCAACCACTGATCCCTCAGAAATGGAGATTTTGCCATGAAGACC 206338
Qy 2379 eutThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuT 2399
Db 206337 TGAAGTATGTGGCCCTTGAGTACCTTAGCTGCCCTGATGATCACTTCCCTTCTCA 206278
Qy 2399 hrlLeuLeuArgGluLeuLeuArgSerLeuGlnHisGlyIleArgArgAsnLeuThrAlaAla 2419
Db 206277 CCTCTCTTGAAGCCCTTGTCTTCCCAACGATGGATGCCAGCAACTCAACAGCTGCCC 206218
Qy 2419 euGlyLeuAlaGluLeuValPheLeuLeuGlyIleAsnGluAlaAspLeuPro----- 2436
Db 206217 TGGGCTGGCCCAAGTGTCTTCTCTGGGATTCACAGGCTGACCT-CCCTGTAAAG 206159
Qy 2436 ----- 2436
Db 206158 TACTTCTCTCCAGAGACTTCCCAACTTCCAGCCCAAGGCTCATGTGTCCCT 206099
Qy 2436 ----- 2436
Db 206098 CTATGACCCCTTGTCCATCCCTGCTCCCAACAGAAATCAATAGTAAAGGCCCATTCAGT 206039
Qy 2436 ----- 2436
Db 206038 AACTGCCCAAGTTGGCCCTTCTGTCAACACAGAGCTTCCCACTCCCTGGGGGAGC 205979
Qy 2437 ----- 2437
Db 205978 TCTTTGATTAAGTGTCTTGGCTCTGTATCATTTGCTGAGCCCTCAGTTTCTTGACAG 205919
Qy 2441 alIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGlu 2461
Db 205918 TCAATGCCATCTCTGTGACTTCTGTACCTTTGACACTTCTCTGGGCTTGTGTGAAG 205859
Qy 2461 lAlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgP 2481
Db 205858 CTTTACACCTGTACCGGGCGGTCAACAGAGTGCAGAGCTCAATGCCAGCCCATGCGTT 205799
Qy 2481 heTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr----- 2494
Db 205798 TCTACTACATGCTGGGGGCTGGGGCTTCTCTGTTCATACACAGTATCTGCTGCTTCC 205739
Qy 2494 ----- 2494
Db 205738 TGACCTTAGAGCTCCCTGTTGTAGAGGCTGGGGGTCACTTCACTGATGTATCCCCAAAC 205679
```

QY	2494	-----	2494	-----
Db	205678	CATAGAAATTCCTGAGAAATCCCTTCCCTTAATCTTGGAACAAGCCTGGATTAGGTGTACAGG	205619	
QY	2494	-----	2494	-----
Db	205618	GCTAGGCTTCTCTTGGAGAGGGGAGTGGACAGAGGCTCTTCTGTATGCTCCCACTTG	205559	
QY	2494	-----	2494	-----
Db	205558	GCTCTGTGAACCTTGTATGATGATCATCATCTTGCTCCACCTTGAGACCTGTCTTGCTGCT	205499	
QY	2495	-----GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAsp	2510	
Db	205498	GTCCTGTGCTCCAGAGTCTGGCTGTGGGCTTGGATCCCAAGGCTATGGAAACCTGACT	205459	
QY	2510	heCyTrpLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPhe	2530	
Db	205438	TCTGTGTGCTCTCCATCTACATACAGTACGCTCATGTGAGTTTGTGTGACAGGTGCTTTG	205379	
QY	2530	IaValSer-----	2532	
Db	205338	CTGTTTCGGTAAAGTCTTGAAGAAACAATTAGGGTAAACAAGATGGACAGGCTCTACAGAT	205319	
QY	2532	-----	2532	-----
Db	205318	TCCCTTTAGAGAAAGCTGAGGCTCTGGGCCCTCGGGCAAGAAAGTAAGTCTCTGC	205259	
QY	2533	-----MetSerValPheLeuTyrIleLeuAlaAla	2543	
Db	205258	TCTCTGTCTCTCATATGTCCCACTGACAGATAGTGTCTTCCGTATATCTGTGTGGCCC	205199	
QY	2543	rgAlaSerCyValAlaIaGlnArgGlnGlyPheGlyuIysIysGlyProVal-----	2559	
Db	205198	GAGCTCTGTGTCTGTCCCACTCAAGGCTTCGAGAAAGAAAGGCCCTGTGTAGTACTGG	205139	
QY	2559	-----	2559	-----
Db	205138	CAGGTGCTGGGACAGGGCGGGTGTGTGTATAGAGAGCATCATGGCGGGTGTGTATATA	205079	
QY	2560	-----SerIleuGlnProSer	2565	
Db	205078	GAGAGCATCATGGCGGAATCATCATATGCCCTTGCCCTCCAGCTCAAGGCTCGGCTCTCC	205019	
QY	2566	PheAlaValLeuLeuLeuLeuSerIleThrTyrLeuLeuAlaLeuLeuSerValIaSer	2585	
Db	205018	TTCACTGTCTCTCTGTCTGTGTGATGTGCATCATGGCTGTCTGGCACTGTGTCTGTCAACAGC	204959	
QY	2586	AspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle-----	2600	
Db	204958	GACACTCTCTCTCTTCACATCACTCTTCGCGGCTGTCAATTTGT-CCAGGTACCTAGCCC	204900	
QY	2600	-----	2600	-----
Db	204899	CGCCTTACTGATGGAGAGAGCATGCTCTGGGGGTGGGCTGTGCTCAACCATGGCTGTCT	204840	
QY	2601	-----GlnGlyProPheIlePheLeuSerTyrValIaLeuSerIysGluValArgIys	2618	
Db	204839	GCTTCCAGGGGCCCTCTCATCTTCTCTCTCTCATAGTGTGTCTGACAAAGGAGTCCGAGAA	204780	
QY	2619	AlaLeuIysLeuAlaCysSerArgIysProSerProAspProAlaLeuThrThrIysSer	2638	
Db	204779	GCACTCAAGTTGGCTCTGCAGCGGAGAACCCAGCCTTGAACCTTGAACACATAGTCT	204720	
QY	2639	ThrLeuThrSer-----	2642	
Db	204719	ACCTGACCTC-GGTAGAGGGGCCGGGCTGTAGCTCAAGAGCGCTTGTGTGAAGGGG	204661	
QY	2642	-----	2642	-----
Db	204660	AAGCAGGCCCCAGAGAACTGTGTGTCTTGTATCTTTTAACTGTGGGCATCTTTTTCCC	204601	

[illegible]

D		203532	TCCCTGGCGAGGAGGGGTGGGAACAGCCTGTGGGTCTTGAGTGAAGAATCTGCCCTTC	203473
OY		2775	iSerrThPrOlyS-----	2779
D		203472	ACAAGTACCCTCAA- GGGTGGGCAGCACAGAGGCTGCAGCCATTGGGGGGGCAACGAAAG	203414
OY		2779	-----	2779
D		203413	CCTAAGAGGCTAAGGCGCTTTGGCTATCCTGTCTACCTCTCTGGGACTCATAGGCTTCC	203354
OY		2780	-----AspGIyGIyProGIyProGIyLyValaProTrPProGIyaAppheGIYT	2796
D		203353	TTTTCTTGAAGANTGGGGGTCCAAGGGTCTGGGAAGTCCCTTGGCCAGGAAGCTTTGGGA	203294
OY		2796	hThralaLyGIyUserSerGIyAsnGIyAlaProGIyLuArgLeuArgIyuAnGIYA	2816
D		203293	CCACACMAAGGAGGAATAGGTGTATGAGGGCCCCCTTGAAGAAGGGGCCACGGGAGAAATGAG	203234
OY		2816	sPalaleuSerArvGIyUGIySerLeuGIyProleuProGIySerSerAlaGINProHSL	2836
D		203233	ATGCTCTACTCGGAAAGGGTCTCTGGAACCCCTTCGGGACCCTTCTACCAACTCACA	203174
OY		2836	ySGIyLle-----	2838
D		203173	AAGGTTA- GTGAGAGGTCTCTCCATCTGCTTCACTCTGCAGATCCTCTTGTGGAAATTTT	203115
OY		2839	-----LeuLySIyLSyCySL	2844
D		203114	AGATGCTCATCTCATGTGTAACCAACCTCTTACCAGAGCATCTCTCAAGAAAGATGTC	203055
OY		2844	eubProthrIIeSerGIyUlySerSerLeuLeuArgLeuProleuGIyCyrThrGIYS	2864
D		203054	TGCCACACATCACACGAAAGATGAGCTCTTAAGGTGTGCTCTTGAGACGGGCAACAGGT	202995
OY		2864	eISerArvGIySerSerAlaSerGIyGIySerArvGIyGIyProProProArvProProp	2884
D		202994	CTTCCCGGGGCTCCACCGCATGAGAGGGCAGACGAAACGACCTCTCCACGCCACAC	202935
OY		2884	roArgInSerLeuGIyUGIyLeuAbnGIyValMetProIIeaImetSerIIelySA	2904
D		202934	CGGCGCAGTCTCTCCAGAAACAGCTGAATGGGGTCAATGCCATGCCAGATCAAG	202875
OY		2904	IaGLyThrValaAPGIyUpSerSerGIySerGIyU	2915
D		202874	CAGGCACTGTGATGAGGACTCTTCTGGCTCCGAG	202840
RESULT 14				
AC095125/c				
LOCUS				
DEFINITION			Rattus norvegicus clone CH230-BN10, WORKING DRAFT SEQUENCE, 2	
ACCESSION			unordered pieces.	
VERSION			AC095125	
KEYWORDS			HTGS_HTSASEL; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE			Rattus norvegicus (Norway rat)	
ORGANISM			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
			Rattus.	
REFERENCE			1 (bases 1 to 254975)	
AUTHORS			Muzny,D.,Marle,, Metzker,M.,Lee,, Abramzon,S., Adams,C., Alder,J.,	
			Allen,C., Allen,H., Alsbrooks,S., Amlin,A., Angilano,D.,	
			Anyaliebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,	
			Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Bernamed,F.,	
			Blewalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,	
			Bryant,N., Buhey,C., Burch,P., Burrell,K., Calderon,E.,	
			Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,	
			Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,	
			Cleveland,K., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,	
			Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,	
			Delgado,O., Denison,S., Deramo,C., Ding,X., Dinh,H., Divya,K.,	
			Draper,H., Dugan-Rocha,S., Dunn,A.C., Durbin,K., Duval,B., Evans,K.,	

	Egan,A., Bascotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabril,A., Ganta,R., Garcia,A., Garner,I., Garza,M., Gebartegorgis,B., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W., Gunatragere,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,J., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulys,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jecod,L., Jiang,H., Johnson,K., Johnson,R., Jolivett,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowals,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshew,L., Louisedge,H., Loraed,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmood,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Mnaj,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaekelemel,O., Okwonu,G., Olamugboon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Plannkoch,C., Plopper,F., Polindebergh,A., Slasson,I., Sitter,C.D., Smay's,D., Snead,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reutter,M., Richards,S., Riggs,F., Rivers,C., Rodery,T., Rojce,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Slason,I., Slight,C.D., Smay's,D., Snead,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steagle,M., Strong,R., Sutton,A., Swalek,A., Taber,P., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valdes,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wlecyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 254975)
AUTHORS	Morley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 254975)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 9, 2003 this sequence version replaced g1:24940723. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/atlas/">http://www.hgsc.bcm.tmc.edu/projects/atlas/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
	----- Genome Center ----- Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project name: GenCNC Center project name: GCNC Center clone name: CH230-8N10 ----- Summary Statistics -----

Assembly program: Atlas;  
 Consensus quality: 231495 bases at least Q40  
 Consensus quality: 232513 bases at least Q30  
 Consensus quality: 233358 bases at least Q20  
 Estimated insert size: 240763; sum-of-coverage estimation  
 Quality coverage: 10x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.bjsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.bjsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 253729: contig of 253729 bp in length  
 \* 253730 253829: gap of unknown length  
 \* 253830 254975: contig of 1146 bp in length.  
 Location/Qualifiers  
 1. 254975  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-8N10"  
 misc\_feature 1. 1872  
 /note="wgs contig"  
 misc\_feature 19784..21080  
 /note="wgs contig"  
 BASE COUNT 60412 a 54867 c 56326 g 62413 t 20957 others  
 ORIGIN

## Alignment Scores:

Prod. No.: 0 Length: 254975  
 Percent Similarity: 11983.00 Matches: 2789  
 Best Local Similarity: 38.64% Conservative: 47  
 Query Match: 77.09% Mismatches: 75  
 DB: 2 Indels: 4437  
 Gaps: 33

US-09-916-849A-3 (1-2923) x AC095125 (1-254975)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 242219 ATGCGGAGCGGGGCTCCAGCGCCCTCCCAACGCCACTGCTGCTACTGCTG 242160  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 242159 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242100  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysProSer 60  
 Db 242099 GGGTCCGGGGGAGCGCGCTCTCTGCGGCTGCGCCCTCGTGGCTGCTGCGCCAGCC 242040  
 QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 242039 TCTGCTTCAACCTCTGCGCTCAACACCGCTGCAAGATTCGGGCACTTGAAGCTGACC 241980  
 QY 81 GlyHisLeuValProHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100  
 Db 241979 GGCACCTGAGTCCCATCAACAGGCTGAGGCTTGTGTCTCAGAAATCAAGGGGTCTCAT 241920  
 QY 101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
 Db 241919 ATCCCTCTCCGCGCATCTCTGAGAGCTGCGCTGAGCTGTCACTCTCGGATACGGA 241860  
 QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnHisProCysIleLeuVal 140  
 Db 241859 GAGCACCTTCTTCCAAAGGCAAGCTGAGCTGCTCAAGAAACCTTTGTTAAAGGCC 241800  
 QY 141 ProArgLeuArgCysGlnSerCysIleValGlnAlaProGlyLeuArgIleGlyIle 160

Db 241799 CCACGGCTCAGATGCCAGTCTTGCAGCTGCGACAGGCCCCAGGGCTCAGGGCTGAGGA 241740  
 QY 161 ArgSerProGlnGlnSerLeuGlnGlyValArgArgValArgAsnValAsnThrAlaProGln 180  
 Db 241739 GGATACAGAGGAAATCTATGGGTGGGCGAGAAAGAAATGAAATACAGCTCCCAAG 241680  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200  
 Db 241679 TTCCAGCTCCCAAGTACCAAGCCAGGCGGCGAGAACCAAGCTGCTGATACCTCTGT 241620  
 QY 201 AlaSerLeuAlaGlnIleAspProAspGlnGlyValGlyValGlyLeuGlyTyrThrMet 220  
 Db 241619 GCTTCTTAAAGGCACTTGAATCCGAGAGGCTAGGGGCTGCACTTGAATCAACCAATG 241560  
 QY 221 AsnAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 241559 GATGCCCTCTTGAATAGCGCTCCAAATCATTTCTTCTTGAACCCCAATCACTGGTCT 241500  
 QY 241 ValThrThrAlaGlnGlnLeuAspArgGlnThrLysSerThrHisValPheArgValThr 260  
 Db 241499 GTCAACCAAGCTGAGAGCTGATCGGGAGCAAGAGCAACCAATGCTTCAGGGTCAAC 241440  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 241439 GCACAGATCATGGTATGCCCGGACGAGTCCCTTGGCCACACTTATCTTATGTACT 241380  
 QY 281 AsnThrAsnAspHisAspProValPheGlnGlnGlnIleValGlySerLeuArgGlu 300  
 Db 241379 GACACCAATGATCAAGACCTGTTTGTAGCAGCAAGAAATCAAGAGAGCTTCAGGGAG 241320  
 QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro 320  
 Db 241319 AACCTGAGGTTGGCTATGAGGTCTTACAGTCAGGGCCACCGATGGAGGCCCTCTCT 241260  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyLysSerProSerGlnValPhe 340  
 Db 241259 AATGCAACATCTGTACCGCTCTCTGAGGGGCTTGAAGGCAACCTTCAAGAGTCTTT 241200  
 QY 341 GluIleAspProAspSerGlyValIleArgThrArgGlyProValAspArgGluGlnVal 360  
 Db 241199 GAGATCATCTCTCTGCTGCGGTATCCGAAACCGCGGCCCTGATGATCGGAAAGAGTG 241140  
 QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 241139 GAATCTTCAAAATGATGAGTGGAGCGCATGATCAAGGGTCCGGACCCAGGCCACGAGGC 241080  
 QY 381 ThrThrAlaValIleLeuSerValGlnAspAsnAspAsnAlaProGlnPheSer 400  
 Db 241079 TCACAGCTGCTGCTCTCTCTGCTGAGATGATATGACAGGCCCTCCAGTTTACG 241020  
 QY 401 GluIleAspArgValValGlnValArgGlnAspValThrProGlyAlaProValLeuArg 420  
 Db 241019 GAGAAAGGCTACGATGCTTCAAGGTGCGGAGAGATGTGACCCAGAGGCCCTGCTTCCGA 240960  
 QY 421 ValThrAlaSerAspAspAspValGlySerAsnAlaValIleIleTyrSerIleMetSer 440  
 Db 240959 GTCAAGGCTTGGCAAGAGCAAGGCGCAAGCCCTGCTGCTGATATGATATGATG 240900  
 QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValIleSer 460  
 Db 240899 GGCACGCTCGGGGACAGTCTTATCTGATCTCAAGCTGAGGCCCTGATGTGGTAGC 240840  
 QY 461 ProLeuAspTyrGlnThrThrLysGlnTyrThrLeuArgValArgAlaGlnAspGlyIle 480  
 Db 240839 CCACCTCGACTATGAGCAACCAAGAAATATCATCGGAGATCCCGGGCCAGAGATGGTGC 240780  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 240779 CGGCTTCACTTCAAGCTTCCGCTGCTATGATACCGGAGGCTCTAGACATCAATGAC 240720  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 Db 240719 AAGGCCCACTTTTGTCAAGACCCCTTTCAGGCGCAAGGCTTATGAGAGTATCCTTTTA 240660



Db 238499 AGACTTGGCTGGCAGGGGAGTGGGTGGAGGGGAGGCAAAATGGTTCTCAGACCCCTG 238440  
QY 1102 ----- 1102  
Db 238439 TCACCTCATGGGACCCCTGTGATCTCAACAGCCTTTGGGAGGTCAAGAGGCTAACAGAG 238380  
QY 1102 ----- 1102  
Db 238379 ACAAGGATTCACCAAGGGGGTCAAGGACCTGCGTCTCTCTCCCACTGTGGTTC 238320  
QY 1102 ----- 1102  
Db 238319 TCAGTAGAGCTGAGAACTGACTTGGACTTTGTCTAGAGCATTCCTTTCCCTCCAGGC 238260  
QY 1102 ----- 1102  
Db 238259 AGAAGAGCTGGCTTGGCTCCCTTCAGAGCTCCAGGAGGAAAGTAGAGTCAGAGCTC 238200  
QY 1102 ----- 1102  
Db 238199 ATTGTCTGTCCAGACAGACTCACTGCGCTGAGTTCTTGTGGGGTCTGGCTTGGG 238140  
QY 1102 ----- 1102  
Db 238139 GAAAGGCTGCACAAAGCCAGGGCAGATGGACTTTTGGGCTCCCAACCCCGTGGCT 238080  
QY 1102 ----- 1102  
Db 238079 TGTCTTGGGGTGTAGTCTTCTCTGTGACCATGATGAGCATGGCTCAGAGAAACAGA 238020  
QY 1102 ----- 1102  
Db 238019 CACGGGTTCCAGATGCACTAGGACTACTGCTTTGAGGGCAGAGACTCCAGGGCCTG 237960  
QY 1102 ----- 1102  
Db 237959 ACTTGTGAGAGAGAAAGATTTCTAGGGGCGTGGAGCTGGCCAGAACCCAGCTGG 237900  
QY 1102 ----- 1102  
Db 237899 GCATTCTGAGCTTGGTCAAGGCTCAGCCCCCAGACAGAGAGACTGGGGAGGGAG 237840  
QY 1102 ----- 1102  
Db 237839 CAGGACTGGGTAGTAGTATCTATCCCTCTGTCAAGCTGTGAGATGAGAGAGAGCGG 237780  
QY 1102 ----- 1102  
Db 237779 GAGGGGAGACTCAAGTACCAAGACCTTCTCTGAGTCTGGAGTTGGGGCTGGAGAGG 237720  
QY 1102 ----- 1102  
Db 237719 AACTGTAAATTTTTCAGGGGGTGAAGCTATGTGGCTACTCAGGCAATCTTACATC 237660  
QY 1102 ----- 1102  
Db 237659 AGGAGGGAAGATGCTTCGCCCTTCTGTCTATCTTACTGTAGACAGACATT 237600  
QY 1102 ----- 1102  
Db 237599 CTTGATCCACCCCTCAGTGGCTGCAAGAACAGGGAAGAGCTTCCAAAGCTATG 237540  
QY 1102 ----- 1102  
Db 237539 GCTTGTCAAGGTGTAAACACTGAGAGCTCAGTTCCCTCCAGAGAGAGAGCTAGGA 237480  
QY 1102 ----- 1102  
Db 237479 GCACTCTCAGCTGAGCTGGCAGAGAGGCAAGAGCGGCTCTCCCTCCGCTTCCA 237420  
QY 1102 ----- 1102

Db 237419 TCCAGGCTTCAAGTGTGAATCCGGCTAGATTCTTAGGGCCCGTGTGAGAGACGCTCT 237360  
QY 1102 ----- 1102  
Db 237359 GCTTAGCCTTGTCTTCTCACCAGGGCAGAGGCCCAAGTCAAAACCTTAGATAGAG 237300  
QY 1102 ----- 1102  
Db 237299 GGAAGGCTGGCCGACAGTGGCCACTGTGGCTGCTTCTTGCCCTTGTCCCTCC 237240  
QY 1102 ----- 1102  
Db 237239 CCATAGCCCTGACCCCTTAGTCAATGATCAAGAGCAACCCAAAGTTCAAGAGGCTGG 237180  
QY 1102 ----- 1102  
Db 237179 TGAGGAGACTGAGGCCCACTGTCCCTTGTCTTCTCTGTGTTCTTGTCTCTCT 237120  
QY 1102 ----- 1102  
Db 237119 TCCCTCCACCCCTGCCATGGGACCTCTGCACTGCTGTCTGTCACTTCTCCAG 237060  
QY 1102 ----- 1102  
Db 237059 CTTCTGTGCTTCTTCTTATGCTTCTTCACTGCTGGCTTTGTCTTCTCTTTC 237000  
QY 1102 ----- 1102  
Db 236999 TCTCTATGTTAGATCCGTCTTCTCTCTCTCACTGTGGCTGCACTGCCAGCTGGG 236940  
QY 1102 ----- 1102  
Db 236939 CTCATGCTCTCGAGCCCAATGATCCGCTGTATCTCTTCAACACAGAGAA 236880  
QY 1102 ----- 1102  
Db 236879 CCTGCCAAGGCTGCTCACTCCCAAGAGAGTCTTATCAAGAAAGCAAGTTG 236820  
QY 1102 ----- 1102  
Db 236819 GGAGGCTGACTGTGCGTCTTCCCTGTGAGCTGGCTGAGGCTTAAACCAGGTGAT 236760  
QY 1102 ----- 1102  
Db 236759 TGCCTGGTCAAGGCTTACCCAGCTCTCTTCTTAAACAATGAGCCACTCTATTC 236700  
QY 1102 ----- 1102  
Db 236699 CCAACTGCTTCCGAAGAAGAGCAATGAGAGAGAGAGCAATGGTTGTGCTG 236640  
QY 1102 ----- 1102  
Db 236639 GGATGGGAGCAGTAGCAGAGGGGAGTGGCCAGATGTGGGCGAGATGGGAGAGAC 236580  
QY 1102 ----- 1102  
Db 236579 AAAGCATGCGGGAGAGGGCAGTCAAGCCTGGGCTGTGAAACACCATGCGGAGTG 236520  
QY 1102 ----- 1102  
Db 236519 TGGCTTCCAGAGCAGAGAGATGGAAGATTCTTGTCTGTGTTCAGAGAGACTGT 236460  
QY 1102 ----- 1102  
Db 236459 TTGTAGAGTCCCTGTGTCTTCTGTCAAGAGGAACATCCATCCAGAAACACTGGGC 236400  
QY 1102 ----- 1102  
Db 236399 TGTAACTTCAAGCTCTTGAAGCCCAAGTCAAGGCTTCTGAAGAGGCAAGAGTCTGATC 236340  
QY 1102 ----- 1102  
Db 236339 ACCATGTGACCCATGGGAGGCTGGAGGGCTTAGAGGGCTGTGACTTGTCTTGGTA 236280



[illegible]

QY	1102	-----	1102	-----	1102
Db	235199	CAGGGGAACAGAGTCAGTGGAAAGACATGTTTCCAGAAAGCATCTGTCACACCCAAAC	235140		
QY	1102	-----	1102	-----	1102
Db	235139	CCGCCACAGCAGGTGAGAGGGGGGTTATCGTGGGGTTGAGGTGTACGGGGCAGAGTC	235080		
QY	1102	-----	1102	-----	1102
Db	235079	TCTCCTGGTCTCTGGGCCCTCTAGAAAAAGAAAGCAAAAGCCGTTTGGACAGG	235020		
QY	1102	-----	1102	-----	1102
Db	235019	GAAATCCAAAGCAGTCGTCAGGGGACAGTGTATATCTAAATAGGGTGAAGGCAAAACAT	234960		
QY	1102	-----	1102	-----	1102
Db	234959	GCAGGCACTTAGTTACCGCTTCATGCCAGAGTGTGGGAAAAACAACAAAGACGGTTT	234900		
QY	1102	-----	1102	-----	1102
Db	234899	CGCAATGTCCTTTTCTTAAAGCAATATAGGTAGCCCCCTCAGAGAAAGACTTGTCCA	234840		
QY	1102	-----	1102	-----	1102
Db	234839	CAGTGAACAGCTCATTTGGCCACCTTATGCGACTAGAAAAAGGTTTCTTGGGTACC	234780		
QY	1102	-----	1102	-----	1102
Db	234779	TTCGGATATGGGCTAGATTTTAGCTGCTTGACCTTCTATTCGCGATGATCTCCTT	234720		
QY	1102	-----	1102	-----	1102
Db	234719	ATGGTCTCCCTCTTGGTCTCTCTCCTCTGGCCCTCCCTCTTCTTAGTACTCTTCTTC	234660		
QY	1102	-----	1102	-----	1102
Db	234659	CTGGTCTCTCTCCCGGGCCCTCCTCTGTACTCGGTTTCTAGTGTCTTACT	234600		
QY	1102	-----	1102	-----	1102
Db	234599	CCTAGTCTCTCTCCTCTGTGTCTCCCTTCTAGTACTTCTCCGTGGTGTCTCC	234540		
QY	1102	-----	1102	-----	1102
Db	234539	TCCTGCAACACTCCCTCCTCTTGCTCCGTTTCTTGACTCCTCTCTGACAGCTTAC	234480		
QY	1102	-----	1102	-----	1102
Db	234479	TCCCTCCTTCGCTTAAGCAGGGCTTGGGCACTGTCTTGCTTACCCCAAGGCTGA	234420		
QY	1103	-----	1103	-----	1103
Db	234419	TGTTGATTGCTGTGCTTCTCTTCAAGTGTGTCCAAGTGTGACAGCCAGGTCTCACT	234360		
QY	1115	uargyalthrlellethraspglmetleuthrisserillethleuarglenglubs	1115		
Db	234359	CCGTTCACCAATCACAGAGATGTCTCACACAGACATCACCTCGCTTGGAAAG	234300		
QY	1135	pmetserprogluarphleuserprobleuenglyleupheilleglalalavalalaa1	1135		
Db	234299	CATGCTCTCAGAACCTTTCTGTCTCCCACTGTGTGGGACTCTTCATTCAAGGCTGTGGCAGC	234240		
QY	1155	aThrleuAlaThrProProAspHisValValPheAsnValGlnArgAspThrAspAl	1175		
Db	234239	CACATTTGGCACACCCCAAGATCAAGTGTGTCTTCAATGTGCAGAGGGAATACTGATGC	234180		
QY	1175	aProglYglYHisIleuAsnValSerIeuSerValGlyglInpProProglYProglYgl	1195		
Db	234179	CCCAAGCCGGCAATATCTCTCAAGTATGGCTGTCAGTGGGCCAGCCCCCAGAACCCGGGGG	234120		
QY	1195	YglYProPheleuProSerGluAspLeuIngluArgLeuYrYleuAsnArgSerIe	1215		



QY	1319	-----	1319	-----	1462	-----	1462
Db	231899	CTCAGCCAGAACCAACCTTATGATATGAGCTGGGAGAGGGGATTCATATTCAAA	231840		230819	ACAGAGAGCTAGAGGGTCTGTCCGTGGCTGTATTTGATGGTGGGAAGGCTGTTC	230760
QY	1319	-----	1319	-----	1462	-----	1462
Db	231839	CTGTAAAGAGGAGCTTCCAGCATCATCTGTGAGGCTCACTGCTGCTTCTCTCC	231780		230759	ACAGGCTGGGAAGAGGAGAGTGTGCTTTCAGAGGCACTTCATCCCTGTGTGCC	230700
QY	1320	--GlyGluHisGlyGluValSerAlaArgSerGlyArgCysThrProGlyValCysLysA	1339		1462	-----	1462
Db	231779	CAGGGAGAGCATGTGTAAGTGAAGTGGCCGCTCAGGCCGTGTGTACTCCAGAGAGTCTGAA	231720		230659	TTGATGTGGGAATATGTGTAAGCGAGAGTGTCCCTGTGGTGTGTCTACCTCTTCTT	230640
QY	1339	ArgGlyGlyThrCysValAlaLeuLeuValGlyGlyPheLysCysAspCysProSerGlyA	1359		1463	-----	1463
Db	231719	ATGGGGGTACCTGTGTCAACCTGTGTGGGAGGCTTCAAAATGTGACTGCCATCCGGGG	231660		230639	TCTCTGTCCACAGCCGCTGTGGTCAAGACGAACTTCACAAAGGCCCATCTGACAGAA	230580
QY	1359	SPheGluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleT	1379		1478	-----	1478
Db	231659	ACTTGAAGAGCCCTTCTGCGAGGTGACACAGCGAGCTTCCAGCCGCTCTTCATCA	231600		230579	GATGGGCTGTGTGTCGTGATGTGCTGTGACACAGGGGTGTGCTGTGCTTCCGAGCTAT	230520
QY	1379	hrPheArgGlyLeuLysArgGlnArgPheHisPheThrLeuAlaLeuSer-----	1394		1498	-----	1498
Db	231599	CCTTCCGGGGCTTGGCCAGCCCTTCCACTTCACTTGGCCCTCTCTGTGATGAGAGGC	231540		230519	GCTGGGCAACTATCTCTGTGTGCTGCCAGGGGACCCAGAGAGGACAGAAAGTAAAGCA	230460
QY	1394	-----	1394		1515	-----	1515
Db	231539	CCTGGGGTTCGAGAGCATTTGAGAGCGTGGGGTACTCTGTGTCCAGTAAAGGGGTG	231480		230459	GATGGGCGGGGAGAGGGGAGAGGTGACAGAGCTAAGGAGAGGGTATGTGGGTCAACAG	230400
QY	1394	-----	1394		1515	-----	1515
Db	231479	GCAGATCACTGTAAGAAAGTACCAAGAGAGTGGCTGACAGAGCATTTGTATGCTTGGC	231420		230399	GATGGGCTCCAGGTCAAAGATGGGGCTCCGAGTAAAGAACTACATGACACCCCT	230340
QY	1394	-----	1394		1515	-----	1515
Db	231419	AGCCGCGCGCTCACTCTCAGAGATCTGCTGGCACTCTCCCATCTTGAAGATGGGC	231360		230339	TCCCTACAATGTGGGAATTAGTCAAGTCTGTGTGGGCTGTGAGGGGAAGGGAGACAT	230280
QY	1394	-----	1394		1515	-----	1515
Db	231359	GGCCAAGATCTTGGGTTCGTGATTCCTGACCTCCAGCCTTACAGATGTCAAGTGTG	231300		230279	CTGTCACTAGGGGTGCTCTGTCTCTCGCGTCTTCTTAAAGGCGCTCATGGCCTC	230220
QY	1394	-----	1394		1515	-----	1515
Db	231299	CTGTGTCTTAAATGCGCCAGGTAGTGAACGCGGGGTGGTCCCGGTTCTCGGCCCGCT	231240		230219	GCTCAAGTCTTCTTCTTCTTGTGCTTATGTGGCAGGGGTATGAGGAAGAACCA	230160
QY	1395	-----	1395		1516	-----	1520
Db	231239	TGACCATCCCTTATACAGGTTTGTCTACCAAGAGAGGTGACGGGCTACTGTCTAACAG	231180		230159	GACTTAAATCTCTGTCCGTGTCTTCTTATATCTCTGTGCCCAAGTCTCTGACCTGA	230100
QY	1408	LYArgPheArgGlyLysHisAspPheValAlaLeuGlyValIleGlnGluGlnValGlnL	1428		1520	hrGlyProLeuLeuLeuGlyGlyValProAlaLeuPProGlyLysPheProValArgMetA	1540
Db	231179	GCGGCTTCAATGAGAGCATGTACTGTAGCTCTGAGGTGTATCCAGAGGACAGTGCAGC	231120		230099	CAGGCGCTTGTGTCTGGGTGGGTGGCCAGATCTGCCAGAGATTTCCCTGTCCGAATGC	230040
QY	1428	euThrPheSerAla-----	1432		1540	rgGlnPheValGlyCysMetArgAsnLeuGlnValAlaPsePArgHisIleAspMetAla	1560
Db	231119	TCACTTCTCTCAGGTGAGGCCAGCGGCTCTGTCTTGTCCCATCTCTAAAGATTTAT	231060		230039	GCGACTTGTGGGTGACATGAAGAACTCCAGGTGAATGCCGAGCATGTGCAGATGGCCG	229980
QY	1432	-----	1432		1560	SPheIleAlaAsnArgGlyThrValProGly-----	1570
Db	231059	GTCTCCAACTCTGACCTCAAGCATGGCGCTCCCTCCCACTGACAGACTGTGAGCC	231000		229979	ACTTCATGCCCAACATGGGACCTGTGCTGG--TATGAGACCTTGGGGGTACGGGCAAGCC	229921
QY	1432	-----	1432		1570	-----	1570
Db	230999	TCCATCTCTAGACATGCCCATCTTCCCTGCACTCTCTCTCTGACCCAGCTTAA	230940		229920	TGGAGCCAAATGCCATGAGATGTGTGTCATGACACGGAATGCCAGCTGTGTGT	229861
QY	1433	-----	1433		1570	-----	1570
Db	230939	CCGGCTCTTTCTTGAAGGAAATCCACACACAGGTGTCTCCATCTGTGCTGAGAGGGT	230880		229860	TCTTCCGCTACTGTGTGTTTCCAGAGCCCTGGGGTCCCTGAATCCCTTCTCTCCG	229801
QY	1448	SerAspGlyGlnTrpHisThrValGlnLeuLysTyrTyrAsnLys-----	1462		1571	-----	1588
Db	230879	AGTATGTGGCAGTGCATACATGTAAGTGAAGTACTCAATATGAGTGTGGTATGGGGGG	230820		229800	TTCTAGGCTGCCCAACAAAGAAAGTGTGTGACACAAACATTTGCATTAATGTGTGCA	229741
QY	1462	-----	1462		1588	hrCysValAsnGlnTrpAspAlaPheSerCysGluCysProLeuGlnPheGlyLysLys	1608

Db	229740	CTGTGTGAACCAAGTGGGAACGATTAGCTGGAGAGTGTCCCTTAGGCTTCCGGGGCAAGA	229681
Qy	1608	ercYalaIgh-----	1611
Db	229680	GCTGTGTCCCAAGGGTAGAGTGGACGCTGCAAGGTTGGGGCTTGGGCTGTTCAGTTGT	229621
Qy	1611	-----	1611
Db	229620	CTGGGATTACGGGCAAGGCGCTTGGGGGAGGAGCTCTGGGGCAGGGCTCCGCAAGTCAGGCTGTG	229561
Qy	1611	-----	1611
Db	229560	GGTGGAAGAGCGTGTCTGGGAGAGTAAGAAAGTAAAGTAAAGAGAGCAAGACCTAAGT	229501
Qy	1611	-----	1611
Db	229500	CAGAGCAGGCGCTGGGACACAAATTGAGGGCGGGCTTGAAACAGTAAATGCTCCGGTGGG	229441
Qy	1611	-----	1611
Db	229440	CGGGGCAATAGCGAGGACCGGGTGTAGGGGAGGAGGAGGGGTCAATCTCATAGAGCTGCC	229381
Qy	1612	-----GlnMet	1613
Db	229380	ACGCTTCTGATCACTCACTGCCCTGGCCCTTGAAATCACTTTTCATCTGCGCTTCCAGAAATG	229321
Qy	1614	AlaAsnProGlnHisPheLeuGlySerSerLeuValAlaTyrPheIleGlyLeuSerLeuPro	1633
Db	229320	GCCAAATCCCAAGCGTTTCTTGGGACAGACCTTGTGGCTGGCATGGCTTCTCTGTGCC	229261
Qy	1634	IleSerGlnProTyrTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValIleu	1653
Db	229260	ATCTCTAGCGCTGGGACACTCAAGCTCAATGTTCCGCAACGCGCAGGACAGATGGCGTCTGG	229201
Qy	1654	LeuGlnAlaIleThrArgGlyArgSerThrIleThrLeu-----	1666
Db	229200	CTGCAAGCGGTCAACAGAGGGGGCGAGCAACATCACTCTGCAAGTGAATGACAGGGCGGCTGG	229141
Qy	1666	-----	1666
Db	229140	GAGTGTGGCAAGCCCGGCTTGGCTTAATTGTGAAGAGCTTCAAGGGGTTTGAACCAAG	229081
Qy	1666	-----	1666
Db	229080	TGGCTGGGCAAGGTTTGGAGAAAGCCAGACATTGGCTGGGCCAATTGGTGGTCAAGTAC	229021
Qy	1666	-----	1666
Db	229020	TGCAACCTATTGGGCTTAAAGAAATATACCTCTCTCTGTGCCCTCAAGCGGGGAGACTG	228961
Qy	1667	-----GlnLeuArgGln	1670
Db	228960	GAAAGAGCTTACGAGGTGTCTACTAATTTTTTTCTTTCCTCTGTGACACTTCCGGGCA	228901
Qy	1671	GlyHisIleValMetLeuSerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGln	1690
Db	228900	GGCCACGTAGTGGTAAAGTGTGAAGGGCAACAGGGCTCCAGGCTCAACCTCGCTGTGGAG	228841
Qy	1691	ProGlyArgAlaAsnAspGlyAspTyrPheHisAlaGlnLeuAlaLeuGlyValaSerGly	1710
Db	228840	CAAGGCGGACCAATATGATGTGACTGGATCAACGCAAGCTGTCTACTGGAGGCTAAGGGG	228781
Qy	1711	GlyProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGluGlyAsnLeu	1730
Db	228780	GGCCCTGTGGCAACGCAATCTGTCTTGTGCTATGGGCAACAGAAAGCGAGGGTAAATCTG	228721
Qy	1731	GlyProArgLeuHisIleGlyLeuHisLeuSerAsnIleThrValGlyTyrIleProGlyPro	1750
Db	228720	GGCCCTGGGCTGCAACGGGCTGCACTGACCAATATGACAGTGTGGGGGAGTTTCCGGGCA	228661
Qy	1751	AlaGlyValAlaAlaArgGlyPheArgGlyGlyS-----	1761

Db	228660	GCACGACATGTGGCCCCCGGGGCTCCGGGGGCTGTTTGACAGGTAAATCTCCGCCGACCT	228660
QY	1762	-----Leu	1762
Db	228600	TCCATCCCATTCCTACCTGAGCCCTGTGTCCACTGAGGGCTAATGCTGCTTATCCCTT	228541
QY	1763	GIINGIValATgValSerAspThrProGIUGIValAsnSerLeuaspProSerHISGly	1782
Db	228540	CAGGGTGTGAGGGTAAAGCGACACACCCAGGGTGTTAGCAAGTCTGATCCGACCGCTGGG	228481
QY	1783	GIuserIleasnValGIUGINGIYCysSerLeuProaspProCysaspSeranProCys	1802
Db	228480	GAGAGCATCAATGTGTGAGCCAGCGCTGTAGCTGGCCAGATCCCTGTGTACTGATCATGT	228421
QY	1803	ProIlaasnSerTYrCysSerAsnaspThrAspSerTYrSerCysSerCysasp	1820
Db	228420	CCACCAACAGCTACTGTGACGACACACTGGGACACTATTTCTGTAGCTGTATCCAGGT	228361
QY	1820	-----	1820
Db	228360	AAGCTGAGGATGCTGGAAATGCGGGCTGAGTGCACAGGGTCACACTCCGTCTAATATGT	228301
QY	1820	-----	1820
Db	228300	AGCAGTTACATCCAGGGTATAGCCACTGTGTGACAGGTTTCAGAGCGGAATAAATAGG	228241
QY	1820	-----	1820
Db	228240	AGTTGTTAGATGGAAGAAAGATGCCAGTAAAGGGCGGGTCTGAACTTTCCGATCCTTCT	228181
QY	1821	---ProGIYTYrTYrGLYAspAsnCyThrAsnValCysAspLeuasnProCysgluHIS	1839
Db	228180	TGTCACAGGTACTATGTGTGACACTGTGTCAAAATGTGTGTGACTGTGAACCATGCGAGCAC	228121
QY	1840	GIuserValCysThrArgIAsProSerIaProHISGlyTYrThrCysgluCyaspProP	1859
Db	228120	CAGTGTGTGTGTATCCCGAAACCAGGTGACACCCACGGCTTACTGTGCAGGTGTTACCA	228061
QY	1860	AsnTYrIleugIYrProTYrCysgluThrArgIle	1870
Db	228060	AATPACTTGGGGCCCTATTGTGTAGACAGGTGA-GGTGGGGCGGGGCTTACTCTGTATCA	228002
QY	1870	-----	1870
Db	228001	CTCAGCCCCGAGAGTCGAGAGTCGGGCTGAGCTGCACCAATGAAGAAGACTATAGTGT	227942
QY	1870	-----	1870
Db	227941	TGAACCAAGCGGGAGAGATGTCAGTGTGTAGAGGGGGCTCATGCTGTGTTGGGTCTTTC	227882
QY	1871	-----AspGIlnProCysPProArgIYrTPTrPGLYHISProThrCysgluYrProCys	1887
Db	227881	TGCAGGATTTGACCAACTTGTGCCCCGTGTGTGTGGGGACACCCCAAGTGTGTCTATGT	227822
QY	1888	AsnCyAspValSerIYsGLYPhaeAspProaspCysAsnIYsThSerGLYluCySHIS	1907
Db	227821	AACCTCGATGTACGAAAGGCTTTGACCCAGCTGACACAAACAAAGTGGCCAGTGCAC	227763
QY	1908	CysIYs-----	1909
Db	227761	TGCAGGTGAGAGACCCGAATGAGCCCCGCTGTGGTACCTGTCTGTAGCTTCTGACC	227702
QY	1909	-----	1909
Db	227701	CTACAGGGGAAGCAGAGAGATGGGCTTCTTCCAGAGCGGAGCAAGCATCTGTGCT	227642
QY	1909	-----	1909
Db	227641	TTTGCTTATACAGGATGGGAGGTGTGTGTTGCCCTTATATCCGGCTGACGGGGATCC	227582
QY	1909	-----	1909
Db	227581	AGTTCAAGTGTGAGTCAGTGAGCCTGTGACTAGCTAATGTGGGCTGTTTGACGCCCATCTT	227522

```
QY 1909 ----- 1909
Db 227521 GAGTATTGGCTTCTCACTGCTATGTCATTACCTGCTGTGAGAGNGTGAATCCTCTCC 227462
QY 1909 ----- 1909
Db 227461 CCACAGAGACCCAGCATAGAGGAGGCTCTCTGAGAGGCCACAGCTTGAAAGCTCTGCG 227402
QY 1910 ----- -GluNH2 1912
Db 227401 TAGCATCCAGAAAGCATCTTGATGCATGACCCCTTCTTCCCTCTAGAGAAATCAAC 227342
QY 1913 TTTATGProProGlySerProThrCysLeuLeuCysAspCysGlyPheProThrGlySerLeu 1932
Db 227341 TACCGACCCCGCAGCAGCCCACTTGCTCTGTGTGACTGTATACCCCACTGGCTTTTG 227282
QY 1933 SerArgValCysAspProGlyLysArgGlyGlnCysProCysLysPheGlyValIleGlyArg 1952
Db 227281 TCCCGAGTCTGTGACCTTAGAGATGGCCAGTGCATGCAAGCTGGAGTCAATGGGCGT 227222
QY 1953 GlnCysAspArgCysAspAsnProPheAlaGlyValThrThrAsnGlyCysGlyIu----- 1970
Db 227221 CAGTGTGATCGCTGTGACAACTTTGTCTGAGGTCCACCAATGGCTGTGAAGGTAGG 227162
QY 1970 ----- 1970
Db 227161 GGGGCTTTAGCACAGTAGATTCCCTCCTCAACATATGCCAAGAGCTTACACTCCAGAG 227102
QY 1970 ----- 1970
Db 227101 GTTAGAATGAGCTGGGCAATGGGTGCGGAGATGTGAGAAAGAGTTGCGGGCATGATA 227042
QY 1971 ----- ValAsnTyRAspSerCysProArg 1978
Db 227041 GAACGCTTACCGGATACCAATGATGTTCTTCTCAATAGTAATTAAGACGCTGGCCAGG 226982
QY 1979 AlaIleGlyAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaIlePro 1998
Db 226981 GGCATAGAGGCTGGGATCTGGGTGCGCCGCGCACGCGGTTTGGCTACCTGCTCTGCCCC 226922
QY 1999 CysProLysGlySerPhe----- 2004
Db 226921 TGCCCCAAGGGCTCCCTTGGTAGGTTTAGATTACAATATGATATGAACTTGAACCTT 226862
QY 2004 ----- 2004
Db 226861 TACTCTCAGAGGAATAAAGATAGCAGCTAGAGCCAGCTTGAGCTCCGGGGCCCCACT 226802
QY 2004 ----- 2004
Db 226801 TTCTGGAATCTATCTCCAGAGACTCAGTGTCTCACTTCACTTCACTCAACCATCA 226742
QY 2004 ----- 2004
Db 226741 ACCTTAGGCTGACCAACGTCCTATATCCAGCCACACAGAGCTGACGAAATCCTGGCT 226682
QY 2004 ----- 2004
Db 226681 GTCCTGAACAGGGCTGGGAGCTAGACTGGTGTGATTGCAATGGGGTGCCTATCTTAGCTAA 226622
QY 2004 ----- 2004
Db 226621 CTACTTAGTACTGAGTTACCCCACTGAAGAGAGCCGTCGTGAGGATGCGTGC 226562
QY 2004 ----- 2004
Db 226561 TGGGTACTATCCCAAGAGATTGAGATTAGTGTGCAATGTTTCATGGGCAAGGAGTAGC 226502
QY 2004 ----- 2004
Db 226501 TGAAGTAGAAGGAGTAGTGTGATGAGCAGGCGTAGTGAATCTTGGGGCGGTGTGGCAG 226442

QY 2004 ----- 2004
Db 226441 AGGTACATAGGCTCCCTTGGCCACTGTACTGATTTTGTGAGATGTTTGGCTGTGA 226382
QY 2004 ----- 2004
Db 226381 GCTCTGTACACTGGGCTTCTCTGTCTCTCCCTTAAGCCCGCCTTCTTGGCTTCTTCTCC 226322
QY 2005 -GlyThrAlaValArgHisCysAspGlyLysArgGlyTyrLeuProPheAsnLeuPheAs 2024
Db 226321 AGGACCTCTGTGGCCACTGTGATGAGCAGACAGGGGCTGGCTCCGCCAATCCTTTCAA 226262
QY 2024 nCysThrSerIleThrPheSerGlyLeuLysGlyPheAla----- 2037
Db 226261 CTGCACGTCACTCACTTCTCAAGACTAAAGGCTTTCT-AAATTGACCCCTTGTCTTCT 226203
QY 2037 ----- 2037
Db 226202 CCTCTTCTTACACCTCTGCTTGTGAAATAGAGCTCCGGGGTTGGGATTTAGCTCAG 226143
QY 2037 ----- 2037
Db 226142 TGGTAGCGCTTGCTAGAGAGCGCAAGCCCTGGGTTGGGTCCAGCTCCGGGAAAAA 226083
QY 2037 ----- 2037
Db 226082 AAGAACCAAAAAAAAAAAAAAAAAAATAGAGCCCGAATCCTGTTTCTGAGGCTCC 226023
QY 2037 ----- 2037
Db 226022 TCTACCTGACAGAGCCACTTCTTCCCACTGATCTTGAGGCCCTTTACTGCTGAGC 225963
QY 2038 ----- GluAr 2039
Db 225962 CCCAGTCCCAACCATCCTGTCTGACCTCTGCACTTCTGCTCTTCATCCAGGCCAGCG 225903
QY 2039 GLeuGlnArgAsnLysSerGlyLeuAspSerGlyArgSerGlnLysLeuAlaLeuLeu 2059
Db 225902 GCTACAGGAAACCAATACAGGCTGTGACTCAGAGCGCTCCCAAGAGGCTGGCCGTCTCT 225843
QY 2059 uArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrG 2079
Db 225842 GCGTAATGGCCACACAGCACACCTCTGCTACTTGGGAGTGAATGTCAAGGTGGCTTACCA 225783
QY 2079 nLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaTh 2099
Db 225782 GCTGGCCACACGACTGCTGGCTCATGAGAGTCCCAAGGGGCTTTGGCTGTCCGCCAC 225723
QY 2099 rGlnAspValHisPheThrGlu----- 2106
Db 225722 ACAGGATGTGACTTCACTGAGGTGGGCTTGAAGTAGGCAATGGGGGCCCATGAGAG 225663
QY 2106 ----- 2106
Db 225662 AGCCTGTTCTGAGATCTTGGGGAGATCTTGAAGGAGATCTTACTGTAGAGAG 225603
QY 2106 ----- 2106
Db 225602 CTGGCCCAACCCCTGTACAGAGGCTCCATGATAGTCTGTAAACAGAGCGCAACAGTT 225543
QY 2106 ----- 2106
Db 225542 GATTAAAGATACCTCTCTGGGGCTGGGATTTAGCTCAGTGTGATAGAGGCTTAACTAG 225483
QY 2106 ----- 2106
Db 225482 GAAAGCAAGGCCCTGGGTTGGTCCCACTCCGAAAAAAGAAAAAAGAGT 225423
QY 2106 ----- 2106
Db 225422 ACCTCTCTTACAGAGTGTCAAGAAATTAGTACTTGAAGTTGAGGGCTGACCTGAGGC 225363
QY 2106 ----- 2106
```

Db 225362 AGAGCCCACTTATTAACCAAGAGCTCAGTCTTGTCAAGGAGCAAAAGAGCAAAAC 225303  
 QY 2106 ----- 2106  
 Db 225302 AGAGCCAGACATGCAGAGACACATGGGTGCGCTGATTAAAGTGTGAAGAAAGTTTAG 225243  
 QY 2106 ----- 2106  
 Db 225242 AGGCCAGTACCCATGCTGGAAGTGAAGATATGATACAGAGGCTTCCAGGGGAGACT 225183  
 QY 2106 ----- 2106  
 Db 225182 GTGGGAGCTGCTAAGGGAGAGACCTAGTGGCGGGTCTGAACAGGACAGGCTGAACCT 225123  
 QY 2107 ----- 2107  
 Db 225122 TGGATAGAGTTGACCAAGACATATGCACTGTGACCATGCCATCTCCAGAAATCTGCT 225063  
 QY 2109 uatgvalglyseralaleuleuapthralaasnlyearghistrgluleuileglnl 2129  
 Db 225062 GAGGTGGGAGGCGGCTCTCTGATGCGGCAATTAAGGCACTGGGAACTGATCCAGCA 225003  
 QY 2129 nfhnglulgylgylthralatpouleuuglnhistrglulalatralsersalaleuul 2149  
 Db 225002 GACAGAGGTGGGAGCTGCTGCTGCTGCTCAGCACTATGAGGCTTATGCCAGTCCCTGGC 224943  
 QY 2149 aglnaametarghisthryleuserprophetrialevalthrproanille----- 2167  
 Db 224942 CAGAAATATCGGACATCTACCTGAGCCCTTCACTATGTCACACCCACATTGGTGA 224883  
 QY 2167 ----- 2167  
 Db 224882 GGTGTGCTGGGCTGGGGGCGGGGTGTGAGAGTCTGCCCATTTCAAGGAGAGCCA 224823  
 QY 2168 ----- 2168  
 Db 224822 GCCGTGCTGGATGATGATCTTGCACCAATTTCTTCTAGTCACTCTGTGGTGGGCT 224763  
 QY 2174 uaplyvglyasaphealaglialyaleuproargtrglualaleuargllyglul 2194  
 Db 224762 GGAATTAAGGGAACCTTGTCTGGACCAAGCTGCCCCGTATGAGGCGTTAGTGGGAGCG 224703  
 QY 2194 npproapbleuugluthrthrvalilleuoprogluservalphearg----- 2210  
 Db 224702 CCTCCAGATCTTGAAGACCAAGTCAATTTGGCAGAGTGTCTTCAAGAGTCAAGCGGG 224643  
 QY 2210 ----- 2210  
 Db 224642 AACGTGATGTGTAAACGCTTGGGGGTGGGACCCATGCAAGTGGGAGGGCCAGCTGAG 224583  
 QY 2211 ----- 2211  
 Db 224582 TGTAGCAGATTTACTTGCAGAGATGCTCCCATGAGTGGAGATGCGAGACCTGTGTAAG 224523  
 QY 2224 laglmguproglugluleuualargharglialarghisproulileuserghllyg 2244  
 Db 224522 CCCAGGAGACTGAGAGACTGCGCGGCGGAGGAGGACCCAGAACTAGTCAAGGGCG 224463  
 QY 2244 lualalalaservalillellyrargthrleuualaglyleuuleuprohisasnlyr 2264  
 Db 224462 AGGAGTGGGCACTGTATCATTTTACCAACGCTGGCTGAGCTGCCCACTAG 224403  
 QY 2264 spproaplyrargserleu----- 2270  
 Db 224402 ACCCAGATTAAGCGTGAAGGTAAAGCGGCTACGACAGGCTTAGAGGTGGGTTAG 224343  
 QY 2270 ----- 2270  
 Db 224342 CAAGGATGTGAAGCATCTCCAGTATTAAGGAGTCTGTGACCGCTTGTGTACTG 224283  
 QY 2271 ----- 2271  
 ArgValProlyrArgProIleleantThrProValIleSerI 2285

Db 224282 TGTACTTCTGTCTATCAGAGTCCCAAGCGCCAGTCAACACACTGTGTAGCA 224223  
 QY 2285 lSerValHisapapbglugluleuuleuproarglaleuaplysProValIthrValG 2305  
 Db 224222 TCAGTGTCCAGATATATAGAGAGTCTTGTCCAGGGGCACTGAGCAAGCAAGTCAAGTGC 224163  
 QY 2305 lnpheargleuleuugluthrthrvalargthrlyeCysValIphetrpaanhis 2325  
 Db 224162 AGTTCGAGCTGTGAGAGAGGAGGAGGAGCCAGCCCATGTGTCTTGTGAAACATTT 224103  
 QY 2325 erlleu----- 2327  
 Db 224102 CGATCTGTAGAGCTGCTGTACCACTTAAAGCCATGACTTGAACCAAGATCCAG 224043  
 QY 2328 ----- 2328  
 Db 224042 CTCGTGGCTGTCACTCTTCTCTCCGTGACTCCGTCTCTCTGCAAGGTCAAGTGC 223983  
 QY 2331 thrglylytyrserlialargglycysglulvalpheargasnlyserhisvalser 2350  
 Db 223982 ACAGGTGGCTGGTCCGCGAGAGGCTGCGAGGTTGTCTCCGTAAAGAGGACATGTAGC 223923  
 QY 2351 CysGlnCysAsnHisMetThrSerPheAlaValIleuMetAspValSerArgArgL 2369  
 Db 223922 TGCAGTGCACATATACAAAGCTTGTGTGTATGATGTGTCCGAGAGAGGTT 223863  
 QY 2369 ----- 2369  
 Db 223862 GGACTCCAGGGGTAGCCCAAGAGCAGGGGTGGGACCAAGTCAAGGGGCTCAAGGCT 223803  
 QY 2370 ----- 2370  
 Db 223802 GCTCTTTTGAACCCCAACATCTACCTCCAGAAATGTGAGATTTTGGCACTGAAGACC 223743  
 QY 2379 leuthrtyrvalaleuuglyvalthrleuualaleuuleuthrPhepheleu 2398  
 Db 223742 CTGAGTATGTGCTTGTGAGACTCTTGTAGCTGTGCTGATATATCACTTCTCTTCC 223683  
 QY 2399 Thrleuuararglleuargserasnlnhisglylileargargasnleuthrala 2418  
 Db 223682 ACCCTCTTCGAGCCCTTGTGCTCCAAACAGATGACATCCAGCGAACCTCACTGCTGC 223623  
 QY 2419 leuglyleualaglnleuvalpheuleuuglylleasnlnalaspLeuPro----- 2436  
 Db 223622 CTGGGCTGGCCCACTGCTTCTCTGGGAGTCAACAGGCTGACCT-CCGTGAG 223564  
 QY 2436 ----- 2436  
 Db 223563 ATACTTCTCTCCAGAGACTTCCCAACTTCAAGCCCAACAGGCTCATAGTCTCCC 223504  
 QY 2436 ----- 2436  
 Db 223503 TGTATGACCTTGTTCATCCCTGCTCCCAAGCATATGATGAAGGCCATTTCCAG 223444  
 QY 2436 ----- 2436  
 Db 223443 TACCTGCCACAGTTGGCTTCTCTCAACACAGAGCTTCCCACTCCCTGGGGGCA 223384  
 QY 2437 ----- 2437  
 Db 223383 CTCCTTATTTGACGTCTGTGCTCTGATCATGTCTGCGCCCTCAAGTTCTTGCACA 223324  
 QY 2441 ValIlealileuuleuHisPheleuTyreCysThrPheSerTrpAlaleuuglu 2460  
 Db 223323 GTCATTTGCATCTGTCTGCACTTCTGTACCTTTCACCTTCTCTGGGCTGTGCTGAG 223264  
 QY 2461 AlaleuHisleuTyralarglaleuthrglulvalagapvalAsnThrGlyPrometArg 2480  
 Db 223263 GCTTACACCTGTATACCGGCGGTCAAGAGGTGCAAGCTAAATGCCAGCCCATGTGT 223204  
 QY 2481 PheTyTyMetLeuGlyTyrglyValProAlaPheIleThr----- 2494  
 Db 223203 TTCTACTATATCTGGGCTGGGGCTTCTGTCTTCAATCAAGATATCTGTCTCTTTC 223144



[illegible][illegible]





QY	445	GLYGLNpheTyLeuapAlaGlnThrGlyAlaLeuaspValValSerProleuaspTyr	464	QY	805	ValAenaspValAaspAspAlaProGlnPheLeuArgaspSerTyrGlnGlySerVal	824
Db	38877	GGAACATTCTATGTGAGTGTCTGAGACTGAGACCTCGATGTGTGAGGCCACTCGACTAT	38818	Db	37797	GTAATGACGTGAATGACAAATGCCCCCAAGTTCTTCAAGATTTCTTACAGGGAGTGTCT	37738
QY	465	GIuThrThryLeuGIuThrThryLeuArgValArgAlaGlnaspGlyValArgProleu	484	QY	825	TyrGluaspValProProPheThrSerValIleuGlnIleSerAlaThrAspArgAspSer	844
Db	38817	GAGACAAACCAAGATATATACAGCGGATCCGGGCCAAGATGTGTGCGGCTCCACTT	38758	Db	37727	TATGAGATGTGCGACCTTCACAGCGTCTCGAGATCTCAGGACCGATGCGCACTCC	37678
QY	485	SerAspValSerGlyLeuValThrValGlnValLeuaspIleAspAspAspAlaProIle	504	QY	845	GlyLeuaspGlyArgValPheTyrThrPheGlnGlyValAspAspGlyAspGlyAspPhe	864
Db	38757	TCCAAAGTCTCCGGTCTAGTAACCGTGCAGGTCTTGAACATCAATCAACACGCCCCATC	38698	Db	37677	GGTCTGAACGGCAGGGTTTCTTACCTTCACAGAGGGAGACGATGAGATGTGTACTTT	37618
QY	505	PheValSerThrProPheGlnAlaThrValIleuGlySerValProleuGlyTyrLeuVal	524	QY	865	IleValGlySerThrSerGlyIleValArgThrLeuArgGlyLeuaspArgGlyLeuVal	884
Db	38697	TTTGTACGACACCCCTTTCCAGGCCACGGTCTTGAAGATGTACTTTAGGCTTACCTGGTT	38638	Db	37617	ATTGAGATCAATCAGGATCGTGGCACACTGGGAGGCTGGATCGTGAAGATGTG	37558
QY	525	LeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeuGlyTyrArgLeu	544	QY	885	AlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProProAlaArgThrPro	904
Db	38637	CTGCATGTCCAAAGCGATTGATGTGATGCGGAGAGAACGCCCGCTCCGAATACAGCTG	38578	Db	37557	GCCCAGTACATCTTGAAGCGTACCGGGTGGACAAAGGGGATGCGCGACGCGCACACC	37498
QY	545	AlaGlyValGlyHisAspPheProPheThrIleAspAspGlyTyrGlyTyrIleSerVal	564	QY	905	MetGlyValThrValThrValLeuaspValAspAspAspProProValPheGlyGlnAsp	924
Db	38577	GCTGGGGTTGGGACGACTTCCCTTACCATTAACACGCGACAGGCTGATCTCGGTG	38518	Db	37497	ATGGAATGACAGTTACTGTACTGATGTGATGATGATCAATCCCCCTCTTTGAACAGGAT	37438
QY	565	AlaAlaGlyLeuAspArgGlyGlyValAspPheTyrSerPheGlyValGlyAlaArgAsp	584	QY	925	GluPheaspValPheValGlyGlyLeuaspSerProIleGlyLeuAlaValAlaArgValThr	944
Db	38517	GGCGGCGAATGTGACCGGAGAAAGGTTGATTTCTACAGCTTCGAGATGAACCGGAGAC	38458	Db	37437	GAGTTTGAATGATTTTGTGGAAGAGAACGCCCATCGACTGGCTGTGCGCGGGTCA	37378
QY	585	HisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuaspValAsn	604	QY	945	AlaThrAspProaspGlyGlyThrAspAlaGlnIleMetTyrGlnIleValGlyGlyAsn	964
Db	38457	CACGGACCCCGACACTCACCGCTCAGCGATGTGATGAACATCCATCCGATGTCAAT	38398	Db	37377	GCCACCGACCCAGAGGAGCAACATGACAGATCATGTACAGATGTGAGGGCAAT	37318
QY	605	AspAspAspProThrPheThrGlnProGlyTyrThrValArgLeuaspGlyAspAlaIle	624	QY	965	IleProGlyValPheGlnLeuaspIlePheSerGlyGlyLeuThrAlaLeuValAspLeu	984
Db	38397	GACACACCCCAACCTTTACGACGCGGAGTACAGGTGCGGTCAACAGAGACGGGGCT	38338	Db	37317	ATCCCTGAGTCTTCACTGACGTGACATCTTCTCGGGCGAGTACAGGCTCGTGAATTTG	37258
QY	625	ValGlyThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleThr	644	QY	985	AspTyrGluaspArgProGlyTyrValLeuValIleGlnAlaThrSerAlaProleuVal	1004
Db	38337	GTGGGACCAAGTGTGTGACGGTGTGACGCGTGGATCGAGATGTCTACAGTGTCAATAC	38278	Db	37257	GACTACAGAGACCGACCGATATATATCTCGTATCCAGGGTCAAGTGTCTCTGTGGTG	37198
QY	645	TyrGlnIleThrSerGlyAsnThrArgAspArgPheSerIleThrSerGlnSerGlyGly	664	QY	1005	SerArgAlaThrValHisValArgLeuLeuaspArgAspAspAspProProValIleuGly	1024
Db	38277	TACAGATACCAACGCGGCAACCCCGCAACCGCTCTCTATACCGCAAGCGGGGGCT	38218	Db	37197	AGCAGGGCTACGTGTCACATGTCGGCTCTCGACCGCATGTATTAACCCCGCATGTGGGC	37138
QY	665	GlyLeuValSerLeuAlaLeuProleuaspTyrTyrLeuGlyAspGlnTyrValIleuAla	684	QY	1025	AspPheGlyIleLeuPheAspAspTyrValThrAspArgSerSerSerPheProGlyGly	1044
Db	38217	GGCCTGGTCTCCCTGACCTTACCGCTGACATAACCTGAGGCGGACGTATGCTGGCTC	38158	Db	37137	AACCTTGAATTTCTTTTCAACAATAATGTCAACAACGCTCGAGAGCTTCTCTGGGGGT	37078
QY	685	ValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAsp	704	QY	1045	AlaIleGlyArgValProAlaHisAspProaspIleSerAspSerLeuThrTyrSerPhe	1064
Db	38157	GTGACTGCTCGGATGGGACAAAGGACGACACAGCTCAATGTGTGTGATGTCACTGAT	38098	Db	37077	GCTATAGGCGCGGTGCTGCCATGACCCCGATATCTCAGACAGCTGACGTACAGCTTT	37018
QY	705	AlaAspThrHisArgProValPheGlnSerSerHisTyrThrValAsnValIleuGlyAsp	724	QY	1065	GlyArgGlyAspGlyLeuSerLeuValLeuLeuAspAlaSerThrGlyGlyLeuValLeu	1084
Db	38097	GCCAAACCCATCGTCCCGTCTTCCAGACTCCCATATACGCTGATGTATATGAAGAC	38038	Db	37017	GAGCGAGAAATGAATCAACGCTGTCTCTCATGTGCTTCCACCGGAGACTGAAGACTG	36958
QY	725	ArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlyAspThrGlyGlyAsn	744	QY	1085	SerArgAlaLeuAspAspAspAspArgProleuGlyValIleMetSerValIleVal-----	1102
Db	38037	CGGCCAGCGGACCAAGTGTGTGATCAATGCTTACAGAGAGGACACAGGGGAGAAAT	37978	Db	36957	AGCGGGCACTGACAAACACCGGCTCTGGAACCATCATGATGTGTGCTGT-CTCAGG	36899
QY	745	AlaArgIleThrTyrPheMetGlyAspSerIleProGlnPheArgIleAspAlaAspThr	764	QY	1102	-----	1102
Db	37977	GCCCGAATCACTACTTATGAGAGACACATCCCTCACTTCGATCGATCGACAGACT	37918	Db	36898	TAAAGAAATGCCAGTATGTGTGGGTGGGAATATGCTTTAGGAGAGTCTTGAAGACC	36839
QY	765	GlyAlaValThrThrGlnAlaGlyLeuAspTyrGlyAspGlnValSerTyrThrIleuAla	784	QY	1102	-----	1102
Db	37917	GGGGCGGTACCAACCGAGCTGAGCTGACATAAGAAATCAAGTGTCTTACACCCCTGGCC	37858	Db	36838	ACTGAGGTGTGGGCTTCCAAAGAGGGGCTGTGAACCAAGCTCTCAACCTGTGATAG	36779
QY	785	IleThrAlaArgAspAspGlyIleProGlnIlySerSerAspThrThrTyrLeuGlyIleLeu	804	QY	1102	-----	1102
Db	37857	ATCACCGCTCGGGAACAATGGATCCCGCAAGTGGGACACAACTATCTGAGATCTCTG	37798	Db	36778	CTCAGAAATTGTGATCAAGGAACCTGGCTGAGAGAAACCCGAGAAATCTTGCAACT	36719
QY				QY	1102	-----	1102

Db	36718	GGCGAGGCCCCCTGCTGCCCCAGGCTGGGCTGCTTGGTGGCTGCCCCCTG	36659
QY	1102	-----	1102
Db	36658	CCTACAGACTGTGACGTGTGGGGAGACGTTGTAGGCTGCCCTGCTGCTGCTGCCA	36539
QY	1102	-----	1102
Db	36598	AGAAATTGTGAAAAAAGGGCGCTGTGAAGCTTGAAGCCCCAGAGCTGGAGAA	36539
QY	1102	-----	1102
Db	36538	GAGGAGGGGAGCTGGAGAGAGAAAGGGAGGGGGAGGGGGTGGAGAACTTGGCT	36479
QY	1102	-----	1102
Db	36478	GGCAGGGGATGGGGTGGAGGGGAGCAAAATCGGTTCTCAGACCCCTGTCACTTCATG	36419
QY	1102	-----	1102
Db	36418	GCACCTGTGATCTCACAGAGCTTGGCAGGTCAAGGGCTTACAGAGACAAGGATTC	36359
QY	1102	-----	1102
Db	36358	CACCAAGGGGTCAGGACCTGCGCTCTCTCTCCCACTGTGTTCTCATGAGCT	36299
QY	1102	-----	1102
Db	36298	GCAGAACTGACTTGTGATTTGTATAGCATTTCCCTTTCCTCAGGAGAAAGCT	36239
QY	1102	-----	1102
Db	36238	GCTTGGCTCCCTTCAGAGCTCCAGGAGATGATGATCATTTGTCTGTGT	36179
QY	1102	-----	1102
Db	36178	CCAGACAGACTCACTGCCCTGAGTTCTTGGGTGGGCTGCTTGGGGAAAGGCTGCA	36119
QY	1102	-----	1102
Db	36118	ACAAAGCAGGGCAGATGAGCTTTTGGGCTCCCCACCCCCGCTGTTTCTTGGGG	36059
QY	1102	-----	1102
Db	36058	TGCTAGTCTTCTCTGTGACATGATGAGCTGCTCAGAGAAACAGACGGGGTTC	35999
QY	1102	-----	1102
Db	35998	GAGTGCATGAGACTAGCTCTTTGAGGGCAGAGACTTCAGGGCCCTGACTTGGTGAG	35939
QY	1102	-----	1102
Db	35938	AGAGAACAGATTTCTAGGGGGCTGGGAGCTGGCAAGAACCCAGCTGGGCAATTCCTG	35879
QY	1102	-----	1102
Db	35878	GCTTGTCAAGGCTCAGGCCCCCAGACAGCAGACTGGGAGGAGAGAGACTGGGT	35819
QY	1102	-----	1102
Db	35818	GATGATATTCTATCCCTCTGTCTGAGCTCTGAGATGAGAGCAGAGGGGAGGAGCA	35759
QY	1102	-----	1102
Db	35758	CTAGTAGCAAGACCTTTCTGAGTCTGGAGTTGGGGCTGGAGAGAGAACTGTTAAGT	35699
QY	1102	-----	1102
Db	35698	TTTTTCAGGGGTGAGACTATGTGGCTTACTCAGGCACTCTTACATCAGGGAGGGAA	35639
QY	1102	-----	1102

---

Db	35638	GATGCTTTGCCCCCTTCTCTGCTCATACTTTATCTGAGACAGACATTCCTGATCCAC	35579
QY	1102	-----	1102
Db	35578	CCCTAGGTGGCTGAAAGACAGAGAGAGAGAGCTTCAAGGATATGCTTGTGAGGC	35519
QY	1102	-----	1102
Db	35518	TGTAAACCACTGAAGCTCAGTTCCCTCCAGAGAGAGAGAGTATGAGACCTCACC	35459
QY	1102	-----	1102
Db	35458	TGAGCTGGCAGAGAGGAGAGAGAGCGCTCTCCCTCCGCTCTCCATCAGAGCTT	35399
QY	1102	-----	1102
Db	35398	CAGTGTGAATCCGGCTAGATTCTTAGGGCCCGGTGTGAGAGACAGCTTGTAGCTTT	35339
QY	1102	-----	1102
Db	35338	GCTTCTCACAGGGACACAGCCCCAGTCAACCTCTATGATTAAGGAGAAAGCTG	35279
QY	1102	-----	1102
Db	35278	GCCGAGCTAGGGCACTGTGGCTGCTTCTGCTTCTGCTCCCTCCCATCAGCCCT	35219
QY	1102	-----	1102
Db	35218	GACCTTAGAGTCACTACCAAGCAACCCAAAGTTTCAAGAGGCTGTGATGAGAGACT	35159
QY	1102	-----	1102
Db	35158	GACGCCACTGTCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	35099
QY	1102	-----	1102
Db	35098	CTGCCATGGGACCTCTCTGACTGCTCTGTCTGTCACTTCCCAAGCTTGTGTCTG	35039
QY	1102	-----	1102
Db	35038	CTTTCCTTATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT	34979
QY	1102	-----	1102
Db	34978	AGATCTGTCTTCT	34919
QY	1102	-----	1102
Db	34918	CTGCAGCCCAATGTCTCCGGCTGTCTATCTCTTCAAGACACAGAGAACTGCCAAG	34859
QY	1102	-----	1102
Db	34858	CTGTCTACCTCCAGACAGAGAGTCTTATTCAGAGAGACAAGTTTGGAGGCTGACT	34799
QY	1102	-----	1102
Db	34798	CTGCGTCTTCCCTGTGTGAGCTGTGAGGCTTACCCAGGTATATGCTGGGTCA	34739
QY	1102	-----	1102
Db	34738	GAGCTTACCCAGCTCTCTCTTTTAAAAAAGATCAGGCACTCTTATTTCCACCTGCT	34679
QY	1102	-----	1102
Db	34678	TCGAGAGAGAGCAATGAGAGAGAGAGAGCAATGGTTGTTGTGGAGATGGGACA	34619
QY	1102	-----	1102
Db	34618	GTAGCAGAGGGGAGGTGGCAAGATGTGGGCGAGATGGGAGCAGCAAGAATGCTC	34559
QY	1102	-----	1102
Db	34558	GGGCAAGGGGCAAGTCAAGGCTGGGCTGGTGAACACCAATGCCGAGTGTGGCTTCC	34499

QY 1102 ----- 1102  
Db 34498 GACAGAGATGGAAGAATTCTTGTGCTTGTGTTCAAGAGACGTTTGTAGACTCC 34439  
QY 1102 ----- 1102  
Db 34438 CTTGTGCTTTCTGTCAAGAGGAACATCCATCCAGAAAACTGGGCTGTAACTTCAG 34379  
QY 1102 ----- 1102  
Db 34378 CTCTGAGCCCAAGTCAGGCTTCTGACAGAGGCAAAAGTCTGATCACTGTGTAC 34319  
QY 1102 ----- 1102  
Db 34318 CCATGGCAGGCTGAGGGGCTTAGAGGGCTGTGACTTGTGTGTGTAACCTGGGCAAA 34259  
QY 1102 ----- 1102  
Db 34258 GCAAGAGAGCAGAGTGAGCAGCATGACAGAGGCTTGTGAGACAAGCAGATCA 34199  
QY 1102 ----- 1102  
Db 34198 CAACAGAGAGTTGAAGGCGAGTAGGCGCAAGCCCAAGGACCACTGAGATGGTACC 34139  
QY 1102 ----- 1102  
Db 34138 TGCCTTCAGTATGCCCAAGCCTTGAAGTGTGACATATACAGAAATAGGGTCCCATATCG 34079  
QY 1102 ----- 1102  
Db 34078 CTTTCTTACTTGTGCTCAGCAGACATGCTCCAGTCCCATCCCATAGATGTCAATAT 34019  
QY 1102 ----- 1102  
Db 34018 TTGCTTGTGATGAGAGTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 33959  
QY 1102 ----- 1102  
Db 33958 CTCTCTCTCTCTGT 33899  
QY 1102 ----- 1102  
Db 33898 TTCTGTGTACTCTGT 33839  
QY 1102 ----- 1102  
Db 33838 AGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33779  
QY 1102 ----- 1102  
Db 33778 AGTTGCTGCGACGTTTAAAGATGCTTAAATGTAAGAGATTCATCAGTGTGTAGT 33719  
QY 1102 ----- 1102  
Db 33718 TTTTAACTAGGAATACCCCTATACCTTTCTAAGCCGTTTCCCTTTCTTGTGCTCCAGT 33659  
QY 1102 ----- 1102  
Db 33658 TTTCAATTTTAACTTGAATCATTTGCCCTGTGTATGAACATTCAGGTGTGATGTAATT 33599  
QY 1102 ----- 1102  
Db 33598 CTGAGCTTAGGTGCGTGGGTGAGTGGGACCGTGCAGATGTGTAGCTGTGTGTAGTG 33539  
QY 1102 ----- 1102  
Db 33538 AGTGCCTGCTAGGGGATATGAGCAGACCTTTGGGTACAATTTGGTGTAGCCAAATGA 33479  
QY 1102 ----- 1102  
Db 33478 ATGATGAATTCCTAAGAGCCCTAGTGCAGAGAGAGAGCAGCCCAAGTGGGCTCAGGG 33419

QY 1102 ----- 1102  
Db 33418 TGAAGCTGCCGCTCTCTTATTTCTACGATCCACCCCAACCCCAACACCT 33359  
QY 1102 ----- 1102  
Db 33358 TTCTCGACTTGGGCACTGGCGGTGTGACACTGGCCGAGTGTGCTGATGATTC 33299  
QY 1102 ----- 1102  
Db 33298 TTGGGTTGGGAGCCCAAGCTGCCAGAGTTGATGAAAGAAAGCTCGGTGTAGAGGTG 33239  
QY 1102 ----- 1102  
Db 33238 TTGCCATGGGCTTAGAGGTTTGCACGAAAGATTATGAGCTGATGCAAGGGAAACC 33179  
QY 1102 ----- 1102  
Db 33178 AAGTCAAGTGAAGACATGTTTCCAGAACGATCTGTCCACACCACCGCACAGCA 33119  
QY 1102 ----- 1102  
Db 33118 CGGTAGAGGCGGGTTATCGTGGGTTGAGGTGACGGGCAAGAGTCTCTGTGTCC 33059  
QY 1102 ----- 1102  
Db 33058 TCTGGGCTCTTAGAAAGAAAGCAAAAGCCGTTTGGAGCAGGGAATCCAAAG 32999  
QY 1102 ----- 1102  
Db 32998 CAGTGCAGGGGACAGTGTATCTGAATAGGTAAAGGAGAAACATGTCAGGCAGTTA 32939  
QY 1102 ----- 1102  
Db 32938 GTTCAAGCTTATGCCAAGAGATGTGGGAAACAAAGAACGGTTTCGATGTCTC 32879  
QY 1102 ----- 1102  
Db 32878 TTTTCCTTAAGACATATGACGTAGCCCTCAAGAGAGACTTGTCCAAATGACACA 32819  
QY 1102 ----- 1102  
Db 32818 GCTCATGGCCACTTTATGCAATCTAAGAAAAAGTTTCTTGGGTACCTTGCAGATAT 32759  
QY 1102 ----- 1102  
Db 32758 GGGCTAGATTTAGCCTGTGACACTTCTATCCGATGATCTCCTTATGTGTGCTCC 32699  
QY 1102 ----- 1102  
Db 32698 CTCTGTGTCTCTCTCTGACCTCTCTCTCTTGTAGTCTCTTCTGTGTGTCTCT 32639  
QY 1102 ----- 1102  
Db 32638 CTCTCCGCGCCCTCCCTTGTACTCGGTTTCTTGTGTGCTCTTACTCTGTAGTCTC 32579  
QY 1102 ----- 1102  
Db 32578 TCTCTCTGTGTCT 32519  
QY 1102 ----- 1102  
Db 32518 ACTCCCTCTCTGTGCTCGGTTTCTTGACTCTCTCTCTGACAGGCTTACTCCCTCTTC 32459  
QY 1102 ----- 1102  
Db 32458 GCTCTAAGCAGGCTTGGCGCATCTGCTTCTGCTTACCCCAAGGCTGATGTGATTCG 32399  
QY 1103 ----- 1118  
Db 32398 TGTGTCTCTCTGAGATGTGTGTGACAGTGTGACAGCCAGTGTCTCCTCTGTGACCC 32339  
QY 1119 IleIleThrAspIleMetLeuThrHisSerIleThrLeuArgLeuIleuAspMetSerPro 1138



Db	32338	ATCATCAACAGACGAGATGCTCAACACACACATCAGCGCTTGAGAAACATGCTCCA	32279	Db	31258	GATGAGGTGGCTTGATCCAGTTAAAGGTGTGCTGCTAGCAACAGGCTTCATATG	31199
Qy	1139	GIuArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAla	1158	Qy	1319	-----	1319
Db	32278	GAAAGCTTTCTGTCCTCCCACTGCTGGGACTCTTCACTTACAGGCTGTGGACCCATTTGGCC	32219	Db	31198	CTGTGTGATCCCTTCATTAACTTGAGCCAGACTTTCCTCCACCGAAGACTGAGGAG	31139
Qy	1159	ThrProPhePheHisValValPheAenValGlnArgAspThrAspAlaProGlyGlyI	1178	Qy	1319	-----	1319
Db	32218	ACACCCCCAGATCAGGTGTGTGCTTCAATGTGCAGAGGGAATCATGATGCCCAAGGCGGC	32159	Db	31138	GCACTGAGAAACAGGGGTTTGAGATCATGGGAGTTCCAGACACTGACCCCGTGAAGCC	31079
Qy	1179	HisIleAenValSerLeuSerValGlyGlnProProGlyProGlyGlyProPro	1198	Qy	1319	-----	1319
Db	32158	CATATCTCAACGTAGAGCTGTGAGTGGCCAGCCCCAGAGACCGGGGTGGGCAACC	32099	Db	31078	TGCAAAATGCTCATTTCTACTCTCCAAATTGTCTGGGCTGAGAGTCTGTACACC	31019
Qy	1199	PheLeuProSerGluAspLeuGlnIuArgLeuIyrLeuAsnArgSerLeuLeuThrAla	1218	Qy	1319	-----	1319
Db	32098	TTTCTACCTTCAGAGAGATCTGAGAGAGCCCTGTACTCAACCGAGCTGCTCACCGCC	32039	Db	31018	TGCCCCACCTGTAGGCTGTGCTTCTCTCACTCTCTTGTCTGTGTTCTTCCAGACT	30959
Qy	1219	IleSerAlaGlnArgValLeuProPheAspAenIleCysLeuArgGluProCysGlu	1238	Qy	1319	-----	1319
Db	32038	ATCTCAGCTCAGCGCGGTCTCCCTTCGACGACAACTTGTGCTCGAGAACCTTGCGAG	31979	Db	30958	CCGAGGCTTCAGATCTTCCAGCGCCAGGCGCACTGGCTTGCAGTGGCGGCAAGACT	30899
Qy	1239	AsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSer	1258	Qy	1319	-----	1319
Db	31978	AATTAATAGCGCTGTGTGTCTGTGCTGTGCTGCACTCTGCGCCCTTATTCGCTTCC	31919	Db	30898	CAGTCTCTCTGTGCACTCTCTGTCTCAGAACTTGGCCAGACTTGTCTGTGCTGC	30839
Qy	1259	SerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProPro	1278	Qy	1319	-----	1319
Db	31918	TCTTGAGTCTCTTCGCGCCCACTCACTGTGCGGGGTCTGCGGTGCTGCGCCACCA	31859	Db	30838	TGTTGTGAGGTGTGGGTGTGGGGCTTAGAGGAAAGGGGGAGNNNNNNCCNNNGANG	30779
Qy	1279	GlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGly	1298	Qy	1319	-----	1319
Db	31858	GGCTTCACAGCGAGCTACTGCGACAGCTGAGTGGACTCTGTACTCAAGACCTTGGA	31799	Db	30778	GANNGGGGGNNNNNNNGNNNNNNNNANNNNNCNGTNNNNNNNNNNNNNNNNNN	30719
Qy	1299	ProHisGlyArgCysArgSerArgGluGlyIyrTyrThrCysLeuCysArgAspGlyIyr	1318	Qy	1319	-----	1319
Db	31798	CCCCATGGGCACTGCGGAGTGGAGGTGTGTACTTACTCTGTGCGGATGGCTAC	31739	Db	30718	NNNCANNN	30659
Qy	1319	Thr-----	1319	Qy	1319	-----	1319
Db	31738	ACGGGTGAGTCTAAGGCGAGGACAAATGGGGGAGCCCTGCAAGGCTGACGCTGGGCCA	31679	Db	30658	NAGNAGNNNNNTTNN	30599
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31678	AGCAAAATCAGACTAATATATGGGCTGCTCATTAATTCCAAGGATATAAATAGATGC	31619	Db	30598	NCNNNGNNNGNNGCCNNNNNNNNNNCNGNNNNNTTNNNNNNNNNNNNNNNNNNNN	30539
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31618	AAACAGCGTCTTATGATCTTACTGCTGTGCAACTGCTGCTTACTAGTAAGTCAGATA	31559	Db	30538	NN	30479
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31558	AGGAACCTTATGTTCACTGTGCAAGCCCAACACCCACACTGGCACATGCCGCTGGCCA	31499	Db	30478	NN	30419
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31498	CAGCCCAAGAAAGCTGTATATGAGTCACTCCGAGCACTGGGTTGTGTCCAGCTT	31439	Db	30418	AATATCTGAATTTTCAAGCCCTGATTTCTGTCTGTCTCAGGCTTAGGCCCCAGTCT	30359
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31438	TACCACTAATATCTGTAAATGAGCGAATTAAGTTAGCTTCTCTTCCACAGTATTTAAG	31379	Db	30358	CTCAGCTGTGTAGGCGGAGAGATCTTCAAGTTAGACCCCAATGTCTGTCTTGGG	30299
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31378	ATGAGATTAACCAATTAATGATAGTCAAAAGTTGCTGTGAGACTCATGTAATTAATTT	31319	Db	30298	GAAAGCTGTCTTACTTCAACACATGACCCCAACTATTAATCAATAGACTGTGAAC	30239
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	31318	GATGAGAAAGAAAGCTGCTTATGTGTGATGTGTGCTTTGTGTGAGAGAGTCTCGA	31259	Db	30238	CTAGTCATCTCTTAGGAAAGAAACTGATCTCTGACCTTGACAGGAACAAGACAGAG	30179
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	30178	TGGAAAGGGGGAAGAACGAGAGTGGGGCGGCGGCTGGAATCTTGACAGGGCCAGGGC	30119	Db	30178	TGGAAAGGGGGAAGAACGAGAGTGGGGCGGCGGCTGGAATCTTGACAGGGCCAGGGC	30119



```

Db 27958 GTCCTAGTGGCAGGGGTATGGGGTAAAGAAACAAGACTTAATTCCTGTCCGTGTC 27899
QY 1516 -----SerLeuAapLeuThrGlyProLeuLeuGlyGly 1527
Db 27898 TTCTTCATACCTTGTGTGTCCTCCAGAGTCTCTGCACTGACAGAGCCCTTGTCTCTGGGTGG 27839
QY 1527 yValProAapLeuProGlySerPheProValArgMetArgGlnPheValGlyCysMetArg 1547
Db 27838 GGTGCAAGATCTGCTCCGAGAGTTCCTGTCCGAATGCGGCACCTTGTGTGGGCTGCATGAA 27779
QY 1547 GAaLeuGlnValAapSerArgHisAlaSerMetAlaSerPheLeuAlaSerArgGlyThr 1567
Db 27778 GAACCTCCAGGTGAAATGCGGCACGTGCAATGCGCCACTTCATGCGCAACAATGGCAC 27719
QY 1567 rValProGly-----GlnLeuArgGlnGlyHisValMetLeuSerValG 1570
Db 27718 TGTGCTGCG-TATGAGAACCTGGGGGTACGGGCAAGGCTGGAGCCAAATGCAATGAGGA 27660
QY 1570 ----- 1570
Db 27659 TCGGTTGTGATGACACGAGGAATGCCAGCCTGCGTGTCTTCCGTCACGTGCTGTGGT 27600
QY 1571 -----CysProAlaLeuGly 1575
Db 27599 TTCCAGAGCCCTGGGGTCCCTGAGTCCCTTCTCTCCGTTTAGGCTGCCCCACCAAGAA 27540
QY 1575 SAaValCysAapSerAaThrCysHisAlaSerGlyThrCysValAaSerGlnTPAAspAl 1595
Db 27539 GAACCTGTGTGACACACACACTTGGCAATATGTGTGGCACCTGTGTGAACAGTGGGAGGC 27480
QY 1555 aPheSerCysGluCysProLeuGlyPheGlyGlySerCysValGln----- 1611
Db 27479 ATTCACTGCGCAGATGTCCTCTAGGCTTCGCGGGCAAGAGTGTGCCAGGTAAGAGTGG 27420
QY 1611 ----- 1611
Db 27419 CAGCTGCCACAGGTTGGGGCTGGGCTCTGTCAATGTCTGGGAATTAAGGCAAGGCTTG 27360
QY 1611 ----- 1611
Db 27359 GGGCAGGCTCTGGGCAAGGCTCCGGAAGTCAAGCTGTGGGTGAAGAGCTGTCTGGGCA 27300
QY 1611 ----- 1611
Db 27299 GAGTAAGAGAGCTAGAGCAGAGACAAGAGACTATGTCAGAGCAGGCGCTGGCACA 27240
QY 1611 ----- 1611
Db 27239 TGTAGGGCGGGCTGGAACAGGTAATGCTCCGTGGGGCGGGCATAGCAGAGACCCGGT 27180
QY 1611 ----- 1611
Db 27179 GTAGAGGAGAGGAGAGGTCATCTCTCATAGAGCTGCCAGCTTCTGACTCCAGCTCT 27120
QY 1612 -----GlnMetAlaSerProGlnHisPheLeuG 1621
Db 27119 GCCCTTGATCACTTTTTCATCTGCTTCCAGAAATGGCCAAATGCCAGCGTTCCTGG 27060
QY 1621 ySerSerLeuValAlaTPHHisGlyLeuSerLeuProIleSerGlnProTyrLeuS 1641
Db 27059 GAGAGAGCCCTTGTGGCTGTGCAATGGCTCTCTCTCCCACTCTCAAGCCCTGGCAGCTCA 27000
QY 1641 eTLeuMetPheArgThrArgGlnAlaSerGlyValIleLeuGlnAlaIleThrArgGlyA 1661
Db 26999 GCTCATGTTCGGCACAAGCCAGGCAATGGCGTCTGTGAGAGCCGTGACACAGAGGGC 26940
QY 1661 rGSetThrIleThrLeu----- 1666
Db 26939 GCAGACCATCACTCTGAGGTGATGACAGGGCGGTGGAGTGTGCAAGCCCGGCTTG 26880
QY 1666 ----- 1666

```

---

```

Db 26879 GCTAATTTGTAGAGCCTTCAGGGGGTTGGACCCAGGTGCTGGGCAAGGTTGGCAGA 26820
QY 1666 ----- 1666
Db 26819 AGCCAGATTTGGCTGGGCCCATTCGGTGGTCAGTACTGCCACTTATTTGGCTTAAAG 26760
QY 1666 ----- 1666
Db 26759 AAATATTCCTCTCCTCTGCGCCCTCAGCGGCAAGACTGGAAGGCTTACTGAGTGTCT 26700
QY 1667 -----GlnLeuArgGlnGlyHisValMetLeuSerValG 1678
Db 26699 ACTAATTTTTTCTTCCCTCTCTGTCATCAGCTTGGGCAAGGCAAGTGTGTAAGTGTGG 26640
QY 1678 lGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyValArgAlaSerProGlyA 1698
Db 26639 AGGCGACAGGGCTCAGGCTCATCTCTGCTGTGAGCCAGGCGGAGCCAGCCAAATGATGTG 26580
QY 1698 sPTPHHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuS 1718
Db 26579 ACTGGCATCAGCACAAGCTGTCACTGGGAAGCTACGGGGGCGCTGGCCACGCCATCTGT 26520
QY 1718 ePheAapTyrGlyGlnGlnArgAlaGlyGlyValaSerLeuGlyProArgLeuHisGlyLeuH 1738
Db 26519 CTTTGACTATGGGCACACAGAGGACAGAGGTATCTGGGGCTCTGCTGTGACAGGGCTGC 26460
QY 1738 lSLeuSerAaThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyP 1758
Db 26459 ACCTGAGCAATATACAGTGGGGGAGTCTCCGGGCCAGCCAGCAAGTGGCCCGTGGCT 26400
QY 1758 hEarGlyCys----- 1761
Db 26399 TCCGGGGCTGTTGCAGGTGAATCTCTCCCGACCTTCATCCCAATCTTACTGGAG 26340
QY 1762 -----LeuGlnGlyValArgValSerAapT 1770
Db 26339 CCTGTTCCACTGAGGCTAATGCTGCTTTATTCCTTCAGGGTGTAGGGTAAAGCAGA 26280
QY 1770 hProGlnGlyValAaSerLeuAapProSerHisGlyGlnSerIleAaValGlnGlnG 1790
Db 26279 CACCCAGAGGTGTAGCAAGTGTGATCCAGCCCGTGGGAGAGATCAATGTGAGCCAG 26220
QY 1790 lYSerSerLeuProAapProCysAapSerAaProCysProAlaAaSerTyrCysSerA 1810
Db 26219 GCTGTACTGCGCAGATCCCTGTGATCTGAATCATGTCCACCAACAGCTTACTGAGCA 26160
QY 1810 sAaPTTPAapSerTyrSerCysSerCysAap----- 1820
Db 26159 AGAAGTGAAGACTAATCTGTGTAGCTGTGATCCAGGTAACTGAGAGATGCTGGAAATG 26100
QY 1820 ----- 1820
Db 26099 GGGGTGGGCGAGGGTCAACACTCCGTCCTAATATAGCAAGTTACATCAGGGTATA 26040
QY 1820 ----- 1820
Db 26039 GCCATCTGTGACAGGGTTCAAGAGCGGAATTAAGAGTGTATAGTGAAGAGATG 25980
QY 1821 -----ProGlyTyrTyrGlnAaP 1827
Db 25979 CCAAGTAAGGGCGGGTCTTAACCTTTCGATCTCTTGTGTCAGGTTACTATAGTGA 25920
QY 1827 sNCysThrAaValCysAapLeuAaProCysGlnHisGlnSerValCysThrArgLysP 1847
Db 25919 ACTGTACAATGTGTGACTTAACCAAGCAGACAGCAAGTGTGTGTATCCGGAABAC 25860
QY 1847 rOserAlaProHisGlyTyrThrCysGluCysProProAaThrTyrLeuGlyProTyrCysG 1867
Db 25859 CCAAGTGAACCCCAAGGCTACATCTGCGAGTGTATCAAAATTAACCTTGGGCGCTTATGTG 25800
QY 1867 luthArgGly----- 1870
Db 25799 AGACCAAGTA-GGTGGGCGGGGCTTACTCTGTATATCACTCAGCCCGGAGAGTGAAGT 25741

```

QY 1870 ----- 1870  
 Db 25740 CGGGCTGAGCTGACGCAATGAAAGACTACATGTGTTAAACAGCGGAGAGATGC 25681  
 QY 1871 ----- AspGlnProCysP 1875  
 Db 25680 AGTGTAGAGAGGGGCTCATGCTGTGTTGGTGTCTTTCTCGAGGATTTGACCACTTGGCC 25621  
 QY 1875 roArgGlyTyrPTrpGlyHisProThrCysGlyProCysAsnGlyAspValSerIleGlyP 1895  
 Db 25620 CCGGTGCTGTGTGGGACACCCACATGTGTCTGACCACTGCAATGTCAAGAAAGGCT 25561  
 QY 1895 heAspProAspCysAsnGlyThrSerGlyGluCysHisCysGlyP 1909  
 Db 25560 TTAAACCAAGCTGCACAAACAAACAGTGGCCAGTGCCTCAAGGTGAGAGACCCGAATG 25501  
 QY 1909 ----- 1909  
 Db 25500 AGCCCCGCTGTGTATCCTGTCTGTAGCTTGTGGCCCTTACAGGGGAGCAGAGAAT 25441  
 QY 1909 ----- 1909  
 Db 25440 GGGCTTCTCTCAGAGAGCGGAGGCAAGCATGTGTCTTGTGCTTATACAGGATGGAG 25381  
 QY 1909 ----- 1909  
 Db 25380 GGTGGGTGTGTCCTTATATCCGGCTGACGGGGATCCAGTTCCAGTGTGACGTGAG 25321  
 QY 1909 ----- 1909  
 Db 25320 CCTGCACTAGCTATGTGGGCTGTGTCAGCCCATCTGTGATTTGGCTTCTCACTGCT 25261  
 QY 1909 ----- 1909  
 Db 25260 ATGTCAATTACCTGTGTGAAGTGTGAATCTCTTCCCAAGAGACCCAGCATAGGGG 25201  
 QY 1909 ----- 1909  
 Db 25200 AGGCTCTCTGTGGGAGGCGCAGCTTGAGCTGTGCTAGCATCCAGAAACATCTT 25141  
 QY 1910 ----- GluAsnHisTyrArgProProGlySerProT 1920  
 Db 25140 GATCCAGTGAACCCCTTCTTCCCTCTTGAAGATCACTACCCAGAGAGGCCCCA 25081  
 QY 1920 hrcCysleuLeuGlyAspCysTyrProThrGlySerLeuSerArgValCysAspProGlnA 1940  
 Db 25080 CTGGCTCTGTGTGACGTGTACCTGAGCTCTTGTGTCCGAGTGTGTGACCCCTGAGG 25021  
 QY 1940 spGlyGlnCysProCysGlyAspProGlyValIleGlyArgGlnCysAspArgCysAspAsp 1960  
 Db 25020 ATGGCCAGTGTCCATGCAAGCTGAGAGTCAATGGGGCTGAGTGTGATTCGTGACAAAC 24961  
 QY 1960 roPhealagluValThrThrAsnGlyCysGlu 1970  
 Db 24960 CTTTGTCTGAGTCAACCAATGTGTGTAAGGTAGGGGCTCTTACAGCAGTAGGT 24901  
 QY 1970 ----- 1970  
 Db 24900 TCCCTCTCAACATATGCAAGAGCTTACACTCAAGAGTTAAGAAATGAGGCTGGCATTTG 24841  
 QY 1970 ----- 1970  
 Db 24840 GGTGCGGATGTGGAAGAGGTTGCGGGGATGATAGAACTGCTTACCGATTAACCATG 24781  
 QY 1971 ----- ValAsnTyrAspSerCysProalaglaIleGluValGlyTlIeTyrP 1986  
 Db 24780 TGTTCCTTCATATGTAATTTATGACGTGCCACAGCGGCATAGAGGCTGGATCTGGT 24721  
 QY 1986 rpProArgThrArgPheGlyLeuProAlaIalaIaProCysProGlySerPhe 2004  
 Db 24720 GGGCCCGCAGCGGTTTGGGTACTGTCTGTGCCCTTCCCAAGGCTCTCTTGGTA 24661

QY 2004 ----- 2004  
 Db 24660 GGTTTTAGATTACATATGATATTTGAACCTTGACCTTACTCTCAGAGAAATAAGATA 24601  
 QY 2004 ----- 2004  
 Db 24600 GCAGCTAGAGCCACAGCTTGAAGCTCCCGGGGCCCACTTGTGATGTCTATCTCCAGAG 24541  
 QY 2004 ----- 2004  
 Db 24540 ACTCAGTGTCTACCTTACCTTCACTCACTCACTCACTCACTCACTCACTCACTCACT 24481  
 QY 2004 ----- 2004  
 Db 24480 TATATCAAGCCCAACAGAGCTGACGAAATTCGTGTCTGTCTGAACAGGCTGGAGCT 24421  
 QY 2004 ----- 2004  
 Db 24420 AGACTGTGATTTGATGAGGGGTGCTTATCTTGAAGCTAACTACTTACTGAGATTACC 24361  
 QY 2004 ----- 2004  
 Db 24360 CACACTGAAGGAGGCGCTGTGAGAGATAGGCGTGTGGTACTATCCAGAGATTCC 24301  
 QY 2004 ----- 2004  
 Db 24300 AGATTAGTGTGACATGTTTCATGTGGGAGGAGATGCTGAGAGTAAAGCACTAGTGA 24241  
 QY 2004 ----- 2004  
 Db 24240 TGAACAGGCGTAGTATATCTTGGGGCGGTGTGTCAGAGGTTACATAGCGTCCCTTGG 24181  
 QY 2004 ----- 2004  
 Db 24180 CCACGTACTGATTTTGTGTGAGATGTTTGGCTCTGAGCTGTGACACTGGGCTTCTC 24121  
 QY 2005 ----- GlyThrAlaValArgHisCys 2011  
 Db 24120 TGTCTCCCTAAGCCCGGCTTCTTGGCTTCTTCTTCCAGGGACTGTGTGGCCACTGT 24061  
 QY 2012 AspGluHisArgGlyTyrIleuProPheAsnLeuPheAsnGlyThrSerIlePhePheSer 2031  
 Db 24060 GATGAGCAGAGGGGCTGTGCTCCGCAAACTTTCACATGCACTGACGTCACTTCTCA 24001  
 QY 2032 GluLeuGlyPheAla 2037  
 Db 24000 GAGCTTAAGGCTTTGT- AAGTGAACCTGTGTCTTCTCTTCTTCAATACCTGTCT 23942  
 QY 2037 ----- 2037  
 Db 23941 TTGAATAATGAGCTCCCGGGTGGGATTTAGCTCAGTGTGAGAGCGTTGCTTAGGA 23882  
 QY 2037 ----- 2037  
 Db 23881 GCGCAAGGCTGTGGTGTGTGTCCAGCTCCGGGAAAAAAGAACCAAAAAA 23822  
 QY 2037 ----- 2037  
 Db 23821 AAAAAAATAGAGCCCGAATCTGTTCCTGAGGCTCTCTCACTGAGAGGCACTTC 23762  
 QY 2037 ----- 2037  
 Db 23761 TTCCCACTGATCTTGTGAGGCTTTTACTGTGAGCCCAAGTCCCACTCTGTCT 23702  
 QY 2038 ----- GluArgLeuGlnArgAsnGly 2046  
 Db 23701 GACCTCTGCACTTGTGTCTCTCACTCAAGGCGGAGGCTTACAGAGAACTACAGGC 23642  
 QY 2047 LeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisThr 2066  
 Db 23641 CTGACTCAGAGAGGCTCCCAAGAGGCTGGGCGCTCTGCGGTATGCCACAGACACACC 23582  
 QY 2067 AlaGlyTyrPheGlySerAspValIyValAlaTyrGlnLeuAlaThrArgLeuLeuAla 2086

```

Db 23581 TCTGGCTACTTGGCGAGTATGTCAAGGTGGGCTTCAAGCTGGCCACAGACTGGCGCT 23522
Qy 2087 HIsGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlu 2106
Db 23521 CATGAGAGTGGCCAGCGGGGCTTTGGGCTGTCCGACACAGAGATGTGACTTCCACTGAG 23462
Qy 2106 ----- 2106
Db 23461 GTGGGGCTTGAAGTAAAGTATGGGGGCCCAAGTGAAGAGCCGTGTTCTTGAGATCTTGG 23402
Qy 2106 ----- 2106
Db 23401 GGGAGATCTTGAAGGAGATACTTCTACTGTAAAGAGCTGCCACCCCTGTGTACAGAG 23342
Qy 2106 ----- 2106
Db 23341 CCTCCATGATAGTCTGTAAACCAAGAGCCGACACAGATTGATTAAAGTACTTCTCTGG 23282
Qy 2106 ----- 2106
Db 23281 GCGTGGGATTTAGCTCAGTGTGAGAGCGCTTACCTAGGAAAGCGAAGCGCTGGGTTGG 23222
Qy 2106 ----- 2106
Db 23221 GTCCCAAGCTCCGAAAAAGAAAAAGAGTACTTCTCTACAGTGTCAAG 23162
Qy 2106 ----- 2106
Db 23161 AATTAGTACTTACGTTTGCATTGAGGGCTGACCTGAGGCAAGCCCACTATTATACCA 23102
Qy 2106 ----- 2106
Db 23101 GAGTCAAGCTTGTGACGAGGACCAAGAGGCAAAAGAGAGAGATGCAAGAGAC 23042
Qy 2106 ----- 2106
Db 23041 ATGGGTGGCCCTGATTAAAGTGAAGAGAGTTTGAAGCCGAGTACCATGTCTGAA 22982
Qy 2106 ----- 2106
Db 22981 GTGACAGATGATACACAGAGGCTTCCAGGGGAGACTGTGGGACTGTAAAGGAGAGA 22922
Qy 2106 ----- 2106
Db 22921 CCTAGTGGCGGGTCTGAACAGGCAAGGCTGAATTGTATGAGTTGACCAAGC 22862
Qy 2107 ----- 2106
Db 22861 ATATGCCACTGTGACATGCGCCATCTCGAATTCGTGAGGTGGGAGCGCCCTCTCG 22802
Qy 2117 AApThrAlaAsnLysArgHisTrpGluLeuIleGlnGlnThrGluGlyValThrAlaTrp 2136
Db 22801 GATGGGCGCATTAAGAGCACTGGGATCTGATCCAGACAGAGAGGTGGCACTGCTGG 22742
Qy 2137 LeuLeuGlnHisTrpGluAlaValAspAlaLeuAlaGlnAsnMetArgHisThrTrp 2156
Db 22741 CTGTCCAGCACTATGAGGCTTATGCTGAGTCCCTGGCCAGAAATATGCGGCAATCTAC 22682
Qy 2157 LeuSerProPheThrIleValThrProAsnIle----- 2167
Db 22681 CTGAGCCCTTCACTATGTCACACCAACATTGGTGAAGTGGTCTGGGCTGGGGGCGG 22622
Qy 2167 ----- 2167
Db 22621 GGGGTGTGAAGAGTCTGCGCATTTCAAAGGCGACGCGTGTGTGTAGTGAATTG 22562
Qy 2168 ----- 2168
Db 22561 CAACCATATTTCTTCTAGTCACTCTCTGTGGGCGCTGTGATTAAGGGGAACTTGTCTGG 22502
Qy 2182 AlaLysLeuProArgTrpGluAlaLeuArgGlyGlnProProAspLeuGlnThrThr 2201

```

---

```

Db 22501 ACCAAGCTGCCCGTTATGAGCGCTTACGTGGGAGCGCCTTCCAGATCTTGAAGCAC 22442
Qy 2202 ValIleLeuProGlnSerValPheArg----- 2210
Db 22441 GTCAATTTGGCCAGAGTCTGTCTCAAGGTCAAGCGGGAGACGTGAGTGTAAAGGCTT 22382
Qy 2211 ----- 2211
Db 22381 GGGGTGGGGCACCCATGCAATGGGGCCAGGGCCAGGTGATGAGCATTTACTTTCGAGA 22322
Qy 2211 uThrProProValValArgProAlaGlyProGlyGluAlaGlnGluProGluGluLeuAl 2231
Db 22321 GATGCTCCCATGGTGTGAGATCTGCAAGACTGTGTGAAGCCGAGAGACTGAGAGCTGGC 22262
Qy 2231 AArgArgGlnArgArgHisProGlnLeuSerGlnGlyGluAlaValAlaSerValIleI 2251
Db 22261 CCGGGGCGAGGAGGACCCAGAACTGAGTCAAGGCGAGGCGAGGCGCAAGTGCATCAT 22202
Qy 2251 eTyArgThrLeuAlaGlyLeuLeuProHisAsnTyAspProAspLysArgSerLeu-- 2270
Db 22201 TTACCAACACGCTGGCTGAGCTGCTGCCCACTACAGATTAAGCTTACCTGAG 22142
Qy 2270 ----- 2270
Db 22141 GTAAAGCGCTACGACAAAGCTTAAAGGTTGGGGTTAGCAAGGAAATGTGAACATCTC 22082
Qy 2271 ----- 2271
Db 22081 CCAAGTATTAAGGTGTCTGTGACGCGCTTGTGTGATCTTCTGTATCAGAGT 22022
Qy 2272 IProLysArgProIleIleAsnThrProValValSerIleSerValHisAspAspGluG 2292
Db 22021 CCCCAG--CGCCAGATCATCAACACCTGTGTAAAGATCATGTCCACAGTATGAGGA 21963
Qy 2292 uLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArgLeuLeuGluThrG 2312
Db 21962 GCTCTTCCAGAGGCACTGTGACAGGCAAGTCAAGTCAAGTCCAGTCTGTGAGAGCGGA 21903
Qy 2312 uGluArgThrLysProIleCysValPheTrpAsnHisSerIleLeu----- 2327
Db 21902 GGAAGCAACAGGCCATCTGTGTCTTGTGAACATTCATCTGTGAGCGTCCCTGTA 21843
Qy 2327 ----- 2327
Db 21842 CCACCTTAAGCCCATGACTTTTGAACCAAGATCCAGGCTCTGCTGCTCATCTTCTT 21783
Qy 2328 ----- 2328
Db 21782 CTTCCTGACTCGGTCTTCTCTCTGCAAGGTCAAGTGGCAAGGTGGTGGTCCGCAAG 21723
Qy 2338 lYCyGluValValPheArgAsnGlnSerHisValSerCyGlnCysAsnHisMetThrS 2358
Db 21722 GCTGCGAGGTGTCTTCTTCCGTACAGAGCCATGTGACTGCACTGCAACCATATGACA 21663
Qy 2358 erPheAlaValLeuMetAspValSerArgArgLys----- 2369
Db 21662 GCTTGTGCTGATGATGATGTGTCGACAGAGGTGTGATCTCCAGGGGTAGCCAGA 21603
Qy 2369 ----- 2369
Db 21602 GCAGCGGTGGGACCAAGTCAAGGGGCTCAGGGCCCTGCTCTTTTGTGACCCCAACAC 21543
Qy 2370 ----- 2370
Db 21542 TCATCCCTCAGAAATGGTGAATTTTGGCACTGAAGACCCGAGATGATGTGGCCCTTGGAG 21483
Qy 2386 aThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArgIleLeuArgS 2406
Db 21482 TCACCTTAGCTGGCCGTGATGATCACTTCTCTCCCAACCTTCTGAGCCCTTGTGCT 21423
Qy 2406 erAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlyLeuValP 2426
Db 21422 CCAACCAAGATGGATCCAGGCACTTCAAGCTGCGCTGGGCTGGGCCCAAGCTGGTCT 21363

```

Qy	2426	heleuLeuGIYIIeaSgIAlaAspLeuPro-----	2436
Db	21362	TTCTCTGGGAAATCAACGAGCTGAACCT-CCCTGTAAAGATATTCTTCTCCAGAGACCT	21304
Qy	2436	-----	2436
Db	21303	TCCCAACCTTCAGGCCACGAAGGCTGAAGGTGTGCCCTTATGACCTTGTCCATCCCT	21244
Qy	2436	-----	2436
Db	21243	GCTCCCAACGACATCACTAGTAGTAAGGGCCCATTCAGTACCTGCCACAGATTGGCTT	21184
Qy	2436	-----	2436
Db	21183	CTCTCACAACCAAGGCTTCCCACTCCCTGGGGGACCTCCTGTATTGACTGTCTTGG	21122
Qy	2437	-----PheAlaCyThrValIleAlaIleLeuLeuHisP	2448
Db	21123	CTCCGTATCATGTGCTGCGCCCTCAGTTTGCCTTGGACAGTCAATTGGCATCTGTGCACT	21064
Qy	2448	heleuTYrLeuCyThrPheSerTrpAlaLeuLeuGIuAlaLeuHisLeuTYrArgAla	2468
Db	21063	TCCGTACCTTTGACCTTCTCCGGGCTCTGCTGAGGCGCTTACACCTGTACCGGCGCG	21004
Qy	2468	euthrGIuValArgAspValAsnThrGlyProMetArgPheTYrTYrMetLeuGIYTrpG	2488
Db	21003	TCACAGAGGGCCGAGACGTCAATGCCAGGCCCAATGCTTCTACTACATGCTGGGCTGGG	20944
Qy	2488	IYValProAlaPheIleThr-----	2494
Db	20943	GCGTTCTGTCTTCATCAGAGGATCCGTGCTCCTTTCTGACCTAGAGCTCCCGTTTG	20884
Qy	2494	-----	2494
Db	20883	CTAGGCTGTGGGTCTACCTTCATGATGATACCCCAACCATAGAAATTCCTGGAATCCC	20824
Qy	2494	-----	2494
Db	20823	TTCCCCCTATACCTTGACAGACGCTGGAATTAGGTGTCAAGGGCTAGGCTTCTCCTTGGAGAG	20764
Qy	2494	-----	2494
Db	20763	GGGAGTGAAGAGAGGTCTCTTCTGTATGCTCCCACTGGGCTCTGTGAACCTTGATGACA	20704
Qy	2495	-----GlyLeuA	2497
Db	20703	TCATACTTGGCTCCACCCCTGAGCACCTGTCTGTGCTGTCTCTGTGCTCCAGGCTCTGG	20644
Qy	2497	IaValGIYLeuAspProGIuGIYTYrGIYAsnProAspPheCySTrPLeuSerIeTYrA	2517
Db	20643	CTGTGGGCTTGGATCCCGAAGGCTATGAGAAACCTTCACTTCTGTGGCTCTCCATCTAGG	20584
Qy	2517	sPThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSer-----	2532
Db	20583	ATAGGCTCATCTGGAGTTTGTCTGACCAAGTGGCTTGTCTGTTCGTTAAGTCTTGAA	20524
Qy	2532	-----	2532
Db	20523	AGCAATTTAAGGTAAAGAGATGCGAGGTCCTACAGATTCCCTTTAGAGAGAGCTGAG	20464
Qy	2532	-----	2532
Db	20463	GCTCTGTGGGCCCCGTGGGCAAGAAAGAACTGTCTGTCTCTGTCTCTCTCATGTGCC	20404
Qy	2533	-----MetSerValPheLeuTYrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnA	2550
Db	20403	CACTCAGAGATAGGTGTCTTCTCTGTATATCCGTGGCCGAGGCTCTGTGTCTGCCAAC	20344
Qy	2550	rgGIuGIYpHeGILuLYAlaGlyProAla-----	2559
Db	20343	GTCAGGGCTTCGAGAAAGAAAGCCCTGTGTGTAGTACTGGCAGGTGCTGGGCAAGGGCG	20284

[illegible]



```
Db 19205 CTGAGGTGAAGCCCGGAAATGGAGAGCTGGGAAAGAGGAAAGACTGCATGTAGAAC 19146
Qy 2684 ----- 2684
Db 19145 AAGCCGGCCAGCTGTAGAGAGCCAGTGTCTCAGACTGCCGCCAACAAGGTCAAGTT 19086
Qy 2684 ----- 2684
Db 19085 TAGAAGCAGTAAGGAGACACAAGTAGAGGCTGGAAATCTCATAAGAGCCAGACTGAC 19026
Qy 2685 ----- 2685
Db 19025 CCTACAGCATAGTCTGTCTTTTAAAGGAGAGTCCAGCTGAACCTGGCCAGGTTCC 18966
Qy 2696 OProglYLeuGlyAspProglYSerLeuPheLeuGlyuGlynaAspGlnInHs----- 2714
Db 18965 CCCTGGCTTAGGGGACCCCAAGTGGCTTATTATGAAAGCCCAAGCCCAACACGAGTGA 18906
Qy 2714 ----- 2714
Db 18905 GACAGAGGCTTAGATGCTTAGAGAGGAGGAGGCTGGCTTTTGTAGAGGCGGCTGG 18846
Qy 2715 ----- 2715
Db 18845 GGCAGGTGGAGCAGTGTGATCCCTCCCTGACCTCCAGATCCGACACGACACTGTACA 18786
Qy 2722 eAspLeuSerLeuGlynaAspAspGlnSerGlySerYrAlaSerThrHisSerSerAsp 2742
Db 18785 GTGACCTGTCCCTGGAGATGACCAAGATGCTTCTTACGCTTCAACCACTCATACACA 18726
Qy 2742 eArgLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2762
Db 18725 GC-----GAGGAAAGAGGAGGAGGAGGCGGCTTCCCTGGCAGAGGAGGCTGGG 18678
Qy 2762 ePLeuLeuLeuGlyProglYrAlaGlynaArgLeuProLeuHisSerThrProlys----- 2779
Db 18677 ACAGCTGTGGGCTGTGAGCTAGAGACTGCCCTTCAAGTACCCCAAGGAGGAGGCTGG 18619
Qy 2779 ----- 2779
Db 18618 CACAGACAGAGGCTGACGCAATGGGGGGGCAACGAAAGCTTAAGAGGCTTCTT 18559
Qy 2780 ----- 2780
Db 18558 GGCATACCTGTCTACCTCTCTGGAGCTCATAGGCTTCTTTCTTTGACAGATGGGGTTC 18499
Qy 2783 rGlyProglYrAlaProThrProglYrAspPheGlyThrThrAlaLysGlnSerSerG 2803
Db 18498 CAGGCTGTGGAGAGTCCCTTGGCCAGAGACTTTGGGACCAACAAGAGAGATAGTG 18439
Qy 2803 LysnglyAlaProglLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2823
Db 18438 GTAGTGGGCCCCCTTGAAGAGGCGGCGGAGAGATGAGATGCTTAACTCGGAAAGGT 18379
Qy 2823 eLeuGlyProLeuProglYSerSerAlaGlnProHisLysGlyLe----- 2838
Db 18378 CTCTGGGACCCCTTCCGGGCTTCTTACCAACTCAAAAGGTTA-GTGAGAGTCTCC 18320
Qy 2838 ----- 2838
Db 18319 TCCATCTGCTTCACTCTGCAATCTTGTGGAATTTAGAAATGCTCCATCTCANGTA 18260
Qy 2839 ----- 2839
Db 18259 CCCACACCTCTTACCCAGGATCTCAAGAGAAAGTGTCTGCCACCATCAGGAAAGA 18200
Qy 2851 eArgLeuLeuArgLeuProLeuGlnGlySerThrGlySerSerArgGlySerSerAlas 2871
Db 18199 GTAGCTCTTAAGGTGGCCCTTGGAGACAGGAGGTCTTCCCGGGGCTTCCACCGCCA 18140
Qy 2871 eArgLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2891
```

```
Db 18139 GTGAGGCGACAGGAAACGAGCTTCTCCAGCCCAACCGCGCCAGTCTTCCAGGAAC 18080
Qy 2891 LysLeuAsngLYValMetProIleAlaMetSerIleLysAlaGlyThrValAspGlnAsp 2911
Db 18079 AGCTGAATGGGGTCAATGCCATGCGCATGAGACATGAAGGACGACATGTGATAGGACT 18020
Qy 2911 eArgGlySerGln 2915
Db 18019 CTTCGTGCTCCGAG 18006
```

Search completed: February 14, 2004, 03:42:07  
Job time : 26841 secs

## SUMMARIES

(without alignments)  
5544.945 Million cell updates/sec

### Scoring table

Searched: 2552756 seqs, 1349719017 residues

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

-MODEL=itrate+g2n. model -DEV=sln
-Q=/cgm2.1/USPRO.spool/US091964814/runat.11022004.14315.20929/app.query.fasta_1.307
-DB=Genesep.191003 -QFMT=fastad -SUFFIX=ring -MINMATCH=0 -1 -LOOFC=0
-LOOBEKT=0 -UNIT5=bits -FASTED=1 -END=-1 -MATRIX=blotsum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pico -NORM=exact -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US091916849 @CGN 1 1 0 @runat.11022004.14315.20929 -NEXP=6 -ICPU=3
-NO MMAP -LARGESUBKEY -NEG SCORES=0 -MAIR -DSPLBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELETE=7

```

Database : N\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed

and is derived by analysis of the total score distribution

## SUMMARIES

Result	Score	Query	Length	DB	ID	Description
No.		Match				
1	15545	100.0	8772	22	AA511678	Human Flamingo cdi
2	15545	100.0	10531	25	AB242868	Human GPCR CELSR2
3	15518.5	99.8	8871	22	AA511677	Human Flamingo cdi
4	15593	99.0	11762	25	ABX34546	Human mdcl cDNA s
5	15279	98.3	9321	22	ABK15177	Human PLEKHA9.1
6	13378.5	86.1	9401	22	ABA08648	Human PLMTG 9 cDN
7	12839.5	82.6	9121	21	AACT74401	Human ORFX ORF195
8	8781.5	56.5	9045	22	AA505832	DNA encoding sever
9	8781.5	56.5	11389	22	AA505835	Human novel cyto
10	8781.5	56.5	11389	25	AB244827	Human GPCR CELSR1
11	8736.5	56.2	9087	22	AA244986	Human NOV7 DNA. I
12	8710	56.0	10195	18	AAAT88330	Mouse receptor ME
13	7836.5	50.4	11365	25	AB243586	Human CELSR3 nucl
14	7815.5	50.3	12348	24	ABQ82327	Human NOV2 encodi
15	7766	50.0	11648	22	AAAD08063	Human extracellular
16	7760.5	49.9	13935	25	AA574583	DNA encoding nove
17	4823	31.0	12791	23	ABL11557	Drosophila melano
18	4766.5	30.7	17282	23	ABL11556	Drosophila melano
19	4584.5	29.5	6791	18	AAT88319	Mouse receptor ME
20	3958	25.5	2332	22	AB319447	Human nervous sys
21	3958	25.5	2332	22	AAK83060	Human immune/h
22	3958	25.5	2332	22	AA531490	Human DNA for a n
23	3958	25.5	2332	22	ABQ66814	Human polynucleo
24	3818	24.6	2391	25	ABT31943	Human breast cance
25	3787	24.4	2603	19	AAV07212	Human calcitonin
26	3769.5	24.2	3312	22	AAF96728	Human late stage c
27	3734	24.0	4152	21	AAAF15924	Human prostate car
28	3701	23.8	2695	24	ABAI90074	Human polynucleo
29	3526	22.7	2077	22	ABAI9448	Human nervous sys
30	3526	22.7	2077	22	AAK83061	Human immune/h
31	3526	22.7	2077	22	AA531491	Human DNA for a n
32	3526	22.7	2077	22	ABQ66815	Human polynucleo
33	2838	18.3	1734	22	ABAI9445	Human nervous sys
34	2425.5	15.6	4874	22	ABAO3134	Human protocadher
35	2078	13.4	3620	25	AB233630	Human secretory p
36	1733.5	11.2	15603	25	ABX56292	Human NOV17a CG9
37	1567.5	10.1	15441	23	AA506029	Drosophila melano
38	1520.5	9.8	13860	23	AA587164	DNA encoding nove
39	1520.5	9.8	14756	24	AAV94753	Human pancreatic c
40	1520.5	9.8	14756	24	AAV94753	Human cadherin (Ca
41	1520.5	9.8	20706	23	ABJ06028	Drosophila melano
42	1511.5	9.7	14785	23	ABJ06090	mRNA for hprt prot
43	1477.5	9.5	14105	23	ABL28581	Drosophila melano
44	1451	9.3	925	22	AAK26671	Human immune/h
45	1451	9.3	925	22	AAK31256	Human cDNA encodi

## ALIGNMENTS

RESULT 1	
AAS11678	
ID	AAS11678 standard; cDNA; 8772 BP.

DT	24-OCT-2001 (first entry)
XX	
DE	Human Flamingo cDNA splice variant

KM obesity; hypotension; hypertension; urinary retention; angina pectoris;  
KM myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
KM benign prostatic hypertrophy; psychiatric disorder; neurological disorder;  
KM manic depression; delirium; dementia; severe mental retardation; ss;  
KM anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
KM signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
KM Plasmid; human; splice variant; g-protein coupled receptor; diabetes;  
KM

KM Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KM antifungal; antiviral; antiprotoczoal; anti-HIV; anorectic; antianorectic;  
 KM antiparkinsonian; cardiatic; cerebroprotective; neuroprotective;  
 KM antidepressant; anticonvulsant; antisense therapy; gene therapy.

OS Homo sapiens.

Key Location/Qualifiers  
 CDS 1..8772  
 FT /\*tag= a  
 FT /product= "Human Flamingo protein #2"

XX WO200161003-A1.

XX 23-AUG-2001.

XX 19-FEB-2001; 2001WO-GB00680.

XX 19-FEB-2000; 2000GB-0004196.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Testa TT;

XX WPI; 2001-502792/55.

XX P-PSDB; AAU07054.

PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -

XX Claim 5; Page 29-33; 66pp; English.

XX The sequence represents a cDNA splice variant which encodes a human  
 CC Flamingo polypeptide. Flamingo is a member of the G-protein coupled  
 CC receptor family, which is involved in signal transduction pathways. By  
 CC screening to identify compounds that stimulate or inhibit the function or  
 CC level of the protein, treatments can be developed for various diseases  
 CC and bacterial, fungal, protozoan and viral infections, including HIV,  
 CC cancer, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's  
 CC disease, acute heart failure, hypotension, hypertension, urinary  
 CC retention, angina pectoris, myocardial infarction, stroke, ulcers,  
 CC allergies and benign prostatic hypertrophy. Also treatable are psychotic  
 CC and neurological disorders such as anxiety, schizophrenia, manic  
 CC depression, delirium, dementia, severe mental retardation, Huntington's  
 CC disease and Gilles de la Tourette's syndrome.

XX Sequence 8772 BP; 1696 A; 2814 C; 2564 G; 1698 T; 0 other;

SO Alignment Scores:

Pred. No.: 0 Length: 8772  
 Score: 15545.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-916-849A-3 (1-2923) x AAS11678 (1-8772)

QY 1 MetArgSerProAlaThnGlyValProLeuProThrProProProLeuLeuLeu 20  
 DB 1 ATGGGAGACCCGGCCAGCGGCTCCCTCCCAAGCGCGCGCGCTGCTCTG 60  
 QY 21 LeuLeuLeuLeuProProProLeuLeuGlyValGlyProCysArgSerLeu 40  
 DB 61 TTGCTGCTGCTGCTGCGCGCGCACTATGGAGAACCAAGTGGGCGCTGCTCTG 120  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrProCysProSer 60  
 DB 121 GGGTCCAGGGGACCAAGGCTCTTTCGGGGCGCTGCCCCCAATGGGCTGCTGCTCAATCC 180  
 QY 61 SerAlaSerAenLeuTyrProLeuTyrThrSerArgCysArgAlaGlyTyrThnGluLeuThr 80

DB 181 TGACGCTCGAACCCTGCTGCTCTACACAGCCGCTGAGGGATCGGGCACTGAGTGA 240  
 QY 81 GYHileuValProHishleuValTyrCysProGluSerGlyAlaHis 100  
 DB 241 GGCACCTGATACCCCAACAGATGAGCTTGGTTCAGAAATCCAGGCCAT 300  
 QY 101 IleProLeuProProAlaProGluGlyCysProTrpSerCysArgLeuLeuGly 120  
 DB 301 ATCCCTTACACACAGCTCTGAGAGCTGCGCTGAGAGCTGCTCTCGGGCATTTGA 360  
 QY 121 GYHileuSerProGlnGlyLysLeuThrLeuProGluGluHisProCysLeuVal 140  
 DB 361 GGCACCTTTCACAGGCGCAAGCTCACATGCCAGAGAGCACCGCTTAAAGGCT 420  
 QY 141 ProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGly 160  
 DB 421 CACGGCTCAGATCCAGTCTGACAGCTGACAGAGCCCGGGCTCAGGGCAAGGAA 480  
 QY 161 ArgSerProGluGluSerLeuGlyArgArgLysArgAsnValAsnThrAlaProGln 180  
 DB 481 AGGTCAACAGAAAGATCCCTGGGTGGCGTCGGAAGAAATGTAATACAGCCCCAG 540  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 541 TTCACAGCCCCCAGCTACAGAGCCAGACGCGAGAACAGCCAGCAGCAGCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyValArgLeuGluTyrThrMet 220  
 DB 601 GCATCCCTAGAGGCCATTCACCCGACAGAGGTGAGAGGATGATCTGAGATACACATG 660  
 QY 221 AsnAlaLeuPheAspSerArgSerArgSerArgPheSerLeuAspProValThrGlyAla 240  
 DB 661 GATGCCCTCTTGAATAGCCGCTCAACAGTTCTTCTCCCTGAGCCAGTCACTGATGCA 720  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluTyrThrLysSerThrHisAlaPheArgValThr 260  
 DB 721 GYACACACAGCCAGGAGGATGATCTGAGACCAAGAACCCACCTTCAGGGTCAG 780  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 781 GGCACAGACCAAGGATGCGCCGACAGATGCCCTGCTACACTCACTCATCTTGTTACT 840  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGluTyrLysGluSerLeuArgLys 300  
 DB 841 GACACCAATGACCATATACCTGTGTCAGACGAGAGATACAAAGAGAGCCTCAGGGAG 900  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB 901 AACCTGAGGTGGCTATAGAGGTCTCACTGTCAAGGCAACGATGTATGCCCTCC 960  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysProSerSerGluValPhe 340  
 DB 961 AATGCCAATATTTCTGATCCGCTGCTGAGAGGGGTCTGGGGGAGCCCTCTGAAGCTTT 1020  
 QY 341 GluThrAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
 DB 1021 GAGATCGACCTCGCTCTGGGGTGAATCCGAGACCGGTGCTGTGATCGGGAGAGGTG 1080  
 QY 361 GluSerTyrGlnLeuThrValGluIleAspArgGlnGlyArgAspProGlyProAspSer 380  
 DB 1081 GAATCTTACAGCTGACGATGAGGACAGTGAACAGAGGTGGAGACCGGGAGCTCTCGAAG 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
 DB 1141 ACCACAGCCCGTGTCTTCTTCTGTGAGAGATGACAAATGATATGCCCCCAAGTTAAT 1200  
 QY 401 GluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 DB 1201 GAAGACGCTATGTGTTCAGAGTGAAGGAGATGTGATCCAGGGGCCCCAGATCTCCAA 1260  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValIleTyrSerIleMetSer 440  
 DB 1261 GTCAAGCTCTCGATCGAGCAAGGGAGCAATGCCGTGTGACTATATGATCATGAGT 1320

QY 441 G1yAsnAlaArgG1yGlnPheTyrLeuAspAlaGlnThrG1yAlaLeuAspValSer 460  
Db 1321 GGCAATGCTGGGGAGAGTTTATCTGGATGCCAGATGAGGCTGGAGTGTGAGC 1380  
QY 461 ProLeuAspTyrG1uThrThrLysG1uTyrThrLeuArgValAlaG1aGlnAspG1yG1y 480  
Db 1381 CCTTTATCACTAAGAGCAACCAAGAGATCACTTACGGGTGCCAGCAAGAGTGTGGC 1440  
QY 481 ArgProLeuSerAsnValSerG1yLeuValThrValGlnValLeuAsp11aAspAsp 500  
Db 1441 GGTCCCCCACTCTTAATGTCCTGGCTGGTGAAGTACAGTCTGGAGATATCAAGAC 1500  
QY 501 AsnAlaPro11ePheValSerThrProPheGln1aThrValLeuG1uSerValProLeu 520  
Db 1501 AATGCCCACTTCTGTCAGCAACCCCTTTCCAGGCTACTGCTGGAGAGGTGCTCCCTTA 1560  
QY 521 G1yTyrLeuValLeuH1sValGlnAla11eAspAlaAspAlaAsp1aArgLeu 540  
Db 1561 GGTACCTGTGTCTTCATGTCATGTCATGATCGCTGATGCTGATGATGCTGCTG 1620  
QY 541 GluTyrArgLeuAlaG1yValG1yH1sAspPheProPheThr11eAsnAspG1yThrG1y 560  
Db 1621 GAATACGCTGTGCTGGAGTGGAGTGAATGCTCCCTTCACTCAATCAATGAGCAAGC 1680  
QY 561 Trp11eSerValAlaAlaG1uLeuAspArgG1uG1uValAspPheTyrSerPheG1yVal 580  
Db 1681 TGGATCTCTGTGCTGTGAATGAGACCGGAGGAGATGTTCTTACAGCTTTGGGATA 1740  
QY 581 GluAlaArgAspH1sG1yThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
Db 1741 GAAGCTGAGAGCCATGACATCCAGCACTCATGCTGCGCAGTGTCAAGTGTCAAGTGTG 1800  
QY 601 LeuAspVal1aAsnAspAsnProThrPheThrGlnProG1uTyrThrValArgLeuAsn 620  
Db 1801 CTGGATTCACAGCAACATCCAACTTTTACCAACAGATACACAGTGGCTCAAT 1860  
QY 621 GluAspAlaAlaValG1yThrSerValValThrValSerAlaValAspArgAspAlaH1s 640  
Db 1861 GAGGATCAGCTGTGGGACCAAGGCTGTGACGCTGTGACGCTGTGACGCTGTGAT 1920  
QY 641 SerVal11eThrTyrGln11eThrSerG1yAsnThrArgAsnArgPheSer11eThrSer 660  
Db 1921 AGGTCACTACCTTACCAATCAACAGTGGCAATCTGAAACCCCTTCTCATCAACAGC 1980  
QY 661 G1nSerG1yG1yLeuValSerLeuAlaLeuProLeuAspTyrLysLeuG1uArgG1n 680  
Db 1981 CAAGTGTGTGTGGCTGTGTATCCCTTGTCCCTGCACTGAGCTAACAATGAGCGGAG 2040  
QY 681 TyrValLeuAlaValThrAlaSerAspG1yThrArgGlnAspThrAlaGln11eValVal 700  
Db 2041 TATGTGTGGCTGTATCCGCTCCGATGGACCTGGGAGGACAGGACAGATTGTGTG 2100  
QY 701 AsnValThrAspAlaAsnThrH1sArgProValPheGlnSerSerH1sTyrThrValAsn 720  
Db 2101 AATGTCAACGACGCAACCACTGCTGTCTTCAAGCTCCCTTACAGTCAAT 2160  
QY 721 ValAsnG1uAspArgProAlaG1yThrThrValValLeu11eSerAlaThrAspG1uAsp 740  
Db 2161 GTTATATAGACCGGCGGCGGAGCAACGCTGTGTGATCAAGCGGCAAGATGAGAC 2220  
QY 741 ThrG1yG1uAsnAlaArgG1eThrTyrPheMetG1uAspSer11eProGlnPheArg11e 760  
Db 2221 ACAGGTAGAGATGCCCATCACTTCACTTCAATGAGAGACAGATCCCGAGTTCCGATC 2280  
QY 761 AspAlaAspThrArg1yAlaValThrThrGlnAlaG1uLeuAspTyrG1uAspGlnValSer 780  
Db 2281 GATGCAACACGGGGGCTGTCAACCAAGCTGAGCTGAGCTGAGCAAGAGTCTT 2340  
QY 781 TyrThrLeuAla11eThrAlaArgAspAsnG1y11eProGlnLysSerAspThrThrTyr 800  
Db 2341 TACACCTGTGGCATTAATGCTGTGGGACATGTGCTTCCCAAGATCCGACACCACTTAC 2400

QY 801 LeuG1u11eLeuValAsnAspVal1aAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db 2401 CTGGAGATCTGTGGTGAACGATGATGACAAATGCCCTTCAAGTCTTCCGACAGCTCTAC 2460  
QY 821 GlnG1ySerValTyrG1uAspValProProPheThrSerValLeuGln11eSerAlaThr 840  
Db 2461 CAGGGCAGTGTCTATAGAGATGTGCCACCTTCACTTACGCTCCGTGAGATCAAGCCACT 2520  
QY 841 AspArgAspSerG1yLeuAsnG1yArgValPheTyrThrPheGlnG1yG1yAspAspG1y 860  
Db 2521 GATGTGATTTCTGACTTAATGACAGGGTCTTCTTCACTTCCAGGAGGCGACAGTGA 2580  
QY 861 AspG1yAspPhe11eValG1uSerThrSerG1y11eValArgThrLeuArgAspLeuAsp 880  
Db 2581 GACGTGACTTATTTGTATGTATGCTCAAGTCAAGATGTGTGACAGCTTACGAGGCTGAT 2640  
QY 881 ArgG1uAsnValAlaGlnTyrValLeuAspAlaTyrAlaValAspLysG1yMetProPro 900  
Db 2641 CGAGAGACGTGGCCCAATATGTCTTGGCGGATATGCAAGTGAAGAGGAGTCCCA 2700  
QY 901 AlaArgThrProMetG1uValThrValThrValLeuAspVal1aAsnAsnProProVal 920  
Db 2701 GCCGCAACCTATGAGAGTGAATGATCTGTGTGATGTGATGATATCAATCCCTGTG 2760  
QY 921 PheG1uG1uAspG1uPheAspValPheValG1uG1uAsnSerPro11eG1yLeuAlaVal 940  
Db 2761 TTTGACAGAGATAGTTTATATGTGTGTGTGAGAGAGAACAGCCCATTTGGGCTAGCCGTG 2820  
QY 941 AlaArgValThrAlaThrAspProAspG1uG1yThrAsnAlaGln11eMetTyrGln11e 960  
Db 2821 GCCCGGATCAAGCCCACTGACCCCGATGAGAGCAACAAAGCCAGATTATGATCAAGATT 2880  
QY 961 ValG1uG1yAsn11eProG1uValPheGlnLeuAsp11ePheSerG1yG1uLeuThrAla 980  
Db 2881 GTGAGGAGCAATCTCTGAGGTCTTCACTGAGCAATCTTCTCCGAGAGCTGACAGCC 2940  
QY 981 LeuValAspLeuAspTyrG1uAspArgProG1uTyrValLeuVal11eGlnAlaThrSer 1000  
Db 2941 CTGTGATCTTATGACTTACAGAGACCGGCTGAGTACGTCTGTGTATCCAGGCAAGTCA 3000  
QY 1001 AlaProLeuValSerArgAlaThrValH1sValArgLeuLeuAspArgAsnAsnPro 1020  
Db 3001 GCTCTGTGTGTAGCGGCTACAGTCAAGTCCGCTCTTGTGACGCAATGCAACCCA 3060  
QY 1021 ProValLeuG1yAsnPheG1u11eLeuPheAsnAspTyrValThrAsnArgSerSer 1040  
Db 3061 CCAAGTCTGGGCACTTTATGATCTTTTTCACACATATGTCAACATGTCTCAAGAC 3120  
QY 1041 PheProG1yG1yAla11eG1yArgValProAlaH1sAspProAsp11eSerAspSerLeu 1060  
Db 3121 TTCCTGTGGGGTGCATTTGGCGAGTACCTGCCCATGACCTGATATCTCAATATGTCTG 3180  
QY 1061 ThrTyrSerPheG1uArgG1yAsnG1uLeuSerLeuValLeuLeuAsnAlaSerThrG1y 1080  
Db 3181 ACTTACAGCTTTGAGGGGAAATGAACTCAAGCTGTGTCTGTCAATCTTCAAGGGGT 3240  
QY 1081 GluLeu11eLeuSerArg1aLeuAspAsnAsnArgProLeuGlnAla11eMetSerVal 1100  
Db 3241 GAGCTAAGCTTAAGCCGCGCACTGACACACACACCCGCTCTGAGGCACTATGAGCGTG 3300  
QY 1101 LeuValSerAspG1yValH1sSerValThrAlaGlnCysAlaLeuArgValThr11e11e 1120  
Db 3301 CTGTGTCAAGCGGCTGATCAACAGCGTACCGCCAGTGGCGGTGTGTGATCATCATC 3360  
QY 1121 ThrAspG1uMetLeuThrH1sSer11eThrLeuAspG1uAspMetSerProG1uArg 1140  
Db 3361 ACCGATGAGATGTCAACCAAGCATCAAGCTGTGCTGTGAAGCATGTCAACCGAGCGC 3420  
QY 1141 PheLeuSerProLeuLeuG1yLeuPhe11eGlnAlaValAla1aThrLeuAlaThrPro 1160  
Db 3421 TTCCTTCAACACTGTGATGCTCTTCACTCAAGCGGTGTGCGCCACGCTGTGCAAGCCA 3480  
QY 1161 ProAspH1sValValPheAsnValGlnArgAspThrAspAlaProG1yG1yH1s11e 1180

```
Db 3481 CCGGACCAAGTGTGCTTCAACGTACAGCGGACACCGAGGCCCGGGGGCCACATC 3540
Qy 1181 LeuAenValSerLeuSerValGlyInProProGlyProGlyGlyProProPheLeu 1200
Db 3541 CTCAACGTGAGCTGTCTGGTGGCCAGCGCCGCGGAGGCGGGCGGCCCTTCCTG 3600
Qy 1201 ProSerGluAspLeuGlnGluArgLeuValLeuAenArgSerLeuThrAlaIleSer 1220
Db 3601 CCTCTGAGGACCTTCAGAGAGGCGCTATACCTCAACCGCAGCTCTCAAGCCATCTCG 3660
Qy 1221 AlaGlnAvalLeuProPheAapAapAenIleCysLeuArgGluProCysGluAenTyr 1240
Db 3661 GCACAGCCCGTGTCTCCCTTCAGACCAACATCTCTGCGGAGCCCTCGACAGAACTAC 3720
Qy 1241 MetArgCysValSerValLeuAarPheAapSerSerAlaProPheAlaIleSerSer 1260
Db 3721 ATGCGCTGTGTGCGGTGCTGCGCTTCAGACTCTCGGCGCCCTTCATGACCTCTCC 3780
Qy 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280
Db 3781 GTGCTCTTCGGCCCATCAACCCGTCCGAGGGGCTGGCTGCGCTGCGCCGCGCTTC 3840
Qy 1281 ThrGlyAapTyrCysGluThrGlyValAapLeuCysTyrSerArgProCysGlyProHis 1300
Db 3841 ACGGTGACTACTGCGAGACCGAGGTGACCTCTGCTACTGCGCGCCCTGTGGCCAC 3900
Qy 1301 GlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAapGlyTyrThrGly 1320
Db 3901 GGGCGCTCCCGACCGCGAGGGCGGCTCAACCTGCTCTGTGCGAGTGGCTACAGGGGT 3960
Qy 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysLeuAenGly 1340
Db 3961 GAGCACTGTGAGGTGAGTGTGCTGCTCAAGCGGTTCCACCCGGGTGTGCAAGATGG 4020
Qy 1341 GlyThrCysValAenLeuLeuValGlyGlyPheLeuCysAapCysProSerGlyAapPhe 1360
Db 4021 GGCACCTGTGTCAACTGTGTGGCGGTTCAAGTGCATGTCCTGCGAGACTTC 4080
Qy 1361 GluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380
Db 4081 GAGAAAGCCCTACTGCGACGTGACCAACGCGAGCTTCCCGCCCACTCTTATCACTTT 4140
Qy 1381 ArgGlyLeuAargGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLeuGlyArg 1400
Db 4141 CGCGGCTGCGCCAGCGTTTCACTTCACTTGGCCCTCTGTTTGCACAAAGAGAGGC 4200
Qy 1401 AapGlyLeuLeuLeuTyrAenGlyAargPheAenGlyHisAapPheValAlaLeuGlu 1420
Db 4201 GACGGGTGTGTGTACATGTGGGTTCAATGAGAAAGCATGACTTGTGGCCCTCGAG 4260
Qy 1421 ValIleGlnGlnValGlnLeuThrPheSerAlaGlyLeuSerThrThrThrValSer 1440
Db 4261 GTGATCCAGAGAGAGTCAAGTCACTTCTCTGAGGGAGTCAACACACGAGTGC 4320
Qy 1441 ProPheValProGlyGlyValSerAapGlyGlnTPrHisThrValGlnLeuLeuTyr 1460
Db 4321 CCAATCGTGCCCGAGAGAGTCAAGTGCAGTGCATACGAGTGCAGTGAATACTAC 4380
Qy 1461 AenLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLysValAla 1480
Db 4381 AATAAGCCACTGTGTGTGACAGAGGCTCCACAGGGCCCATACAGACCAAGGTGCT 4440
Qy 1481 ValValThrValAapGlyCysAapThrGlyValAlaLeuAargPheGlySerValLeuGly 1500
Db 4441 GTGGAGACGTGTGAGTGTGACACAGAGTGGCTTGGCTTCGATCTGTCTGGGC 4500
Qy 1501 AenTyrSerCysAlaAlaGlnGlyThrGlnGlySerLysLysSerLeuAapLeuThr 1520
Db 4501 AACTACTCTGTGTGCTGCGACAGGCGACCAAGGTGTGACAAAGAGTCTGTGATCTGACG 4560
Qy 1521 GlyProLeuLeuGlyGlyValProAapLeuProGlySerPheProValArgMetArg 1540
Db 4561 GGGCCCTGCTACTAGGCGGGGTGCTGACCTGCGGAGAGCTTCCAGTCCGAATGCGG 4620
Qy 1541 GlnPheValGlyCysMetAargAenLeuGlnValAapSerArgHisIleAapMetAlaAap 1560
Db 4621 CAGTTCGTGGAGCTGCAAGCGGAACCTGCAAGTGTGACAGCGGACATATGACTGTGAC 4680
Qy 1561 PheIleAlaAenAenGlyThrValProGlyCysProAlaLysLysAenValCysAapSer 1580
Db 4681 TTCACTTCCCAACATGAGCACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4740
Qy 1581 AenThrCysHisAenGlyGlyThrCysValAenGlnTPrAapAlaPheSerCysGlyCys 1600
Db 4741 AACTGTCACCAATGTGGGGCACTTGTGTGAACCAAGTGGACGGGTTCACTGTGAGTGC 4800
Qy 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlyMetAlaAenProGlnHisPheLeu 1620
Db 4801 CCTTGGGCTTGGGGGCAAGCTGTGGCCAGGAATGTGCATATCCAGACACTTCTTG 4860
Qy 1621 GlySerSerLeuValAlaTPrHisGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640
Db 4861 GCAGAGAGCTGTGTGCTGTGACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920
Qy 1641 SerLeuMetPheAargThrAargGlnAlaAapGlyValLeuLeuGlnAlaIleThrAargGly 1660
Db 4921 AGCTCATGTTCGCGACGCGCCAGCGGTGTCTGTGCGAGGCTATCACAGGGGG 4980
Qy 1661 ArgSerThrIleThrLeuGlnLeuAargGlyGlyHisValMetLeuSerValGluGlyThr 1680
Db 4981 GCAGACCACTACCTTACAGTACAGAGAGGCGCATGTAGTGTGAGCGTGTGAGGGGACA 5040
Qy 1681 GlyLeuGlnAlaSerSerLeuAargLeuGluProGlyAargAlaAenAapGlyAapTPrHis 1700
Db 5041 GGGCTTCAAGCTCTCTCTCTCCGTGAGACCGAGGCGGCGCAATGACGTGTGCTGAC 5100
Qy 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaAlaLeuSerPheAap 1720
Db 5101 CATGCAAGCTGTGCACTGTGGAGCCAGCGGGGGGCTGTGGCCATCTGTCTCTTCAT 5160
Qy 1721 TyrGlyGlnGlnAargAlaGluGlyAenLeuGlyProAargLeuHisGlyLeuHisIleSer 1740
Db 5161 TATGTGAGAGAGAGAGAGAGGCGACCTGTGGCCCGCGCTGTGATGTGTGACCTGAGC 5220
Qy 1741 AenIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaAargGlyPheAargGly 1760
Db 5221 AACTATACAGTGGGCGGATATACCTGTGGCCAGCGGCGGTGTGCGCTTCCGGGCG 5280
Qy 1761 CysLeuGlnGlyValAargValSerAapThrProGluGlyValAenSerLeuAapProSer 1780
Db 5281 TGTGTGAGGGGTGTGCGGTGAGCGATACCGCAAGGGGGTTTAAACAGCTGTGATCCAGC 5340
Qy 1781 HisGlyGlySerIleAenValGlnGlnGlyCysSerLeuProAapProCysAapSerAen 1800
Db 5341 CATGGGAGAGATCAACGTGTGAGAGAGCTGTGAGCTGTGACCTTGTACTGAAC 5400
Qy 1801 ProCysProAlaAenSerTyrCysSerAenAapTPrAapSerTyrSerCysAap 1820
Db 5401 CCGTGTCTCTCTAACAGCTATGTCAGACAGACTGTGGACAGTATCTTCACTGTGAT 5460
Qy 1821 ProGlyTyrTyrGlyLysAapAenCysThrAenValCysAapLeuAapProCysGlyHisGln 1840
Db 5461 CAGGTACTATGTGTGACAACTGTACTAATGTGTGTGACTGTAAACCGGTGTGAGCACAG 5520
Qy 1841 SerValCysThrAargLysProSerAlaProHisGlyTyrThrCysGluCysProProAen 1860
Db 5521 TCTGTGTGTACCGGCAAGGCCATGTGCCCCCAAGGCTATACCTGTGAGTGTCCCAAT 5580
Qy 1861 TyrLeuGlyProTyrCysGluThrAargLysAapGlnProCysProAargGlyTPrTPrGly 1880
Db 5581 TACCTTGGGCAATATCTGTGACACAGATTTGACACAGCTTGTGCTCCGTGTGTGGGGA 5640
Qy 1881 HisProThrCysGlyProCysAenCysAapValSerLeuGlyPheAapProAapCysAen 1900
Db 5641 CATCCACATGTGTGCGCATGTGACCTGTGATGTGAGCAAGAGCTTGTAAACCAAGTGCAC 5700
```

QY 1901 LyeThrSerGlyGlyGlySerHisGlyGlyGlyLeuAsnHisGlyThrArgProProGlySerProThr 1320  
DB 5701 AAGACAGACGGCAGGCTGCACTGCAGAGAGAACCACTACGAGCCCCCAGCAGACCCCAACC 5760  
QY 1921 CysLeuLeuCySerProCySerProThrGlySerLeuSerArgValCysAspProGlyAsp 1340  
DB 5761 TGCTCTGCTGTAAGTACCTGACCCCAAGGCTCTGTCGACAGAGCTGAGACCTGAGAGAT 5820  
QY 1941 GlyGlnCysProCysGlyProGlyValIleGlyArgGlnCysAspArgCysAspAspPro 1360  
DB 5821 GGGCAGAGTGCATGACAGCAGGAGTGCATGGGGGTGATGATGATGATGATGATGATGATGAT 5880  
QY 1961 PheAlaGlyValIleThrThrAsnGlyCysGlyValAsnTyThrSerCysProAlaIle 1380  
DB 5881 TTGGCTGAGGTGCACCAACCAATGCTGTGAGAGTATGATGATGATGATGATGATGATGAT 5940  
QY 1981 GluAlaGlyIleTyrTyrProArgTyrThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
DB 5941 GAGGCTGGGATCTGGTGGCCCCGTAACCGCTTGAGGCTGCTGCTGCTGCTGCTGCTGCTG 6000  
QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlyHisArgGlyTyrLeuProPro 2020  
DB 6001 AAGAGCTCTTTGGAGCTGCTGGCGCCACTGTGATGATGATGATGATGATGATGATGATGAT 6060  
QY 2021 AsnLeuPheAsnGlyThrSerIleThrPheSerGlyLeuGlyGlyPheAlaGlyLeu 2040  
DB 6061 AACCTCTTCAACGACGTCATCATCACTTCTCAGAACGTGAGGCTTGGCTGAGCGGCTA 6120  
QY 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuLeuLeuLeuArg 2060  
DB 6121 CAGCGATATAGTCAAGGCTTACAGCTGAGGCGCTCCAGAGAGTACGCTGCTGCTGCTGCTG 6180  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIleValAlaIleArgGlnLeu 2080  
DB 6181 AAGCGACGACGACACACAGCTGGCTACTTGGCAGGAGGAGCTCAAGGCTGCTCAAGCTG 6240  
QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
DB 6241 GCCACGGGCTGTGGCTGCACAGAGACACACAGGAGGCTTGGGCTGTGCTGCACACAG 6300  
QY 2101 AspValHisPheThrGlnLeuLeuLeuLeuArgValIleSerAlaLeuLeuAspThrAlaAsn 2120  
DB 6301 GAGGTGACCTTCACTGAGATCTGCTGGGGTGGGAGCGCCCTCTGAGACACAGCCAC 6360  
QY 2121 LysArgHisTyrGlyLeuLeuIleGlnGlnThrGlnGlyGlyTyrAlaIleThrLeuLeuGlnHis 2140  
DB 6361 AAGCGGCACTGGAGCTGATCCAGACAGACAGAGGTGGCACCGCTGGCTGCTCCAGCAC 6420  
QY 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisGlyThrTyrLeuSerProPhe 2160  
DB 6421 TATGAGGCTTACCGCAGTGCCTGGCCAGACATGGGACACCTAACCCCTTTC 6480  
QY 2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspIleGlyAsnPheAla 2180  
DB 6481 ACCATGTCACCGCCCAACATTTGCACTCCGTAAGTGGCTTGGACAAAGGAACTTTGCT 6540  
QY 2181 GlyAlaLeuLeuProArgTyrGlyAlaLeuArgGlyGlyGlnProProAspLeuGlnThr 2200  
DB 6541 GGGGCGCAAGCTGCCGCTACAGAGGCTGCTGTGGGAGAGCGCCCGGACCTTGAAGACA 6600  
QY 2201 ThrValIleLeuProGlySerValPheArgGlyThrProProValIleArgProAlaGly 2220  
DB 6601 ACGATCATTTGCTGATGCTGTCTTCAAGAGACCGCCCGTGTGTACAGGCCCGGAGGC 6660  
QY 2221 ProGlyGlyAlaGlnGlyProGlyGlyLeuAlaArgArgGlyAspArgHisPProGlyLeu 2240  
DB 6661 CCGGAGAGGCGCCAGAGAGCAGAGAGCTGGCACGGGACAGAGAGCGCACCGGAGGCTG 6720  
QY 2241 SerGlnGlyGlyAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
DB 6721 AGCCAGGAGAGGCTGTGGCCAGGCTCATCATCATCATCATCATCATCATCATCATCATCAT 6780

QY 2261 HisAsnTyrAspProAspIleArgSerLeuAlaGlyValProLysArgProIleIleAsnThr 2280  
DB 6781 CATTAATATGACCTTGCAGACAGGCGAGCTTGAGAGTCCCAAGCGCCGATCATACACACA 6840  
QY 2281 ProValIleSerIleSerValHisAspAspGlyGlyLeuLeuProArgAlaLeuAspIle 2300  
DB 6841 CCGGTGTAGATCAAGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6900  
QY 2301 ProValThrValGlnPheArgLeuLeuGlnThrGlyGlyValIleArgProIleCysVal 2320  
DB 6901 CCGGTACAGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6960  
QY 2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGly 2340  
DB 6961 TTCTGAGACATTCATCACTTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 7020  
QY 2341 ValIlePheArgAsnGlySerHisGlyValSerGlyGlnCysAsnHisMetThrSerPheAla 2360  
DB 7021 GTGTCTTCCGAAAGAGAGCCAGCTCACTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 7080  
QY 2361 ValLeuMetAspValSerArgArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 2380  
DB 7081 GTGTCTATGAGAGTCTTCTGCGGAGAGATGGAGAGATCTGCACTGAGACACTGACA 7140  
QY 2381 TyrValAlaLeuGlyValIleThrLeuAlaLeuLeuLeuThrPhePhePheLeuThrLeu 2400  
DB 7141 TACGTGGCTTACGTGTGCTCACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7200  
QY 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleGly 2420  
DB 7201 TTGCTATCTCGCTTCCACCAACAGCGCATCCAGACGTGACAGCTGACCTGCTGCTGCTG 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
DB 7261 CTGGCTCAGCTGTGTCTTCTCTCGGGAAATCAACAGGCTGACCTGCTTGTGCTGACA 7320  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGly 2460  
DB 7321 GTCATTTGCCATCTGTGTGCTGATCTTGTACCTGTGACCTTGTGCTGCTGCTGCTGCTG 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValIleAspAspValAsnThrGlyProMetArg 2480  
DB 7381 GCTTGTACCTTATACCGGCACTCAAGAGTGCAGATGTACACACCGGCCCATATGCC 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
DB 7441 TTCTACTACATGCTGGGCTGGGCGTGGCTGCTTCACTACAGGCTGAGCGTGGCGCTG 7500  
QY 2501 AspProGlyGlyTyrGlyAsnProAspPheCysTyrPleuSerIleTyrAspThrLeuIle 2520  
DB 7501 GACCCCGAGGGCTACGGAGAACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7560  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
DB 7561 TGGAGTTTGTGCTGGCCCGGTGGCTTTGCGCTTGTGAGAGTGTCTTCTGTACATCTG 7620  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyLeuGlyProValSer 2560  
DB 7621 GGGGCCCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7680  
QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
DB 7681 GGGCTGAGGCTCTCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7740  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
DB 7741 CTCTCTGTCAACAGGACACCTCTCTTCACTACCTTGTGCTGCTGCTGCTGCTGCTGCTG 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerGlyValArgGlyAlaLeu 2620  
DB 7801 CAGGCGCCCTTCACTTCTCTCTCTATGTGGCTTACAGAGAGTCCGGAAGACACTC 7860  
QY 2621 LysLeuAlaCysSerArgIleProSerProAspProAlaLeuThrThrIleSerThrLeu 2640





QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db ATCGAGAGCGCGGCGACGGCGGTCCCTCCCAACGCGCGCGCGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysAspSerLeu 40  
 Db TTGCTGCTGCTGCTGCGCGGCGACCTATGAGAGAACCAAGAGGAGGCTCTGCTCTTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTrpLeuCysProSer 60  
 Db GGGCTCCAGGGGAGCAAGGCTCTTGCGGGGCGCTGGGCCCCCAATGGGCTGGCTCTGTCATCC 242  
 QY 61 SerAlaSerLeuLeuLeuTrpLeuTrpThrSerArgCysArgAspAlaGlyThrGlyLeuThr 80  
 Db TCGAGCTCGAAGCTCTGGCTCTACACCAAGCGCTGCAAGGATCGGGCACTGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisAlaAspGlyLeuArgValTrpCysProGlySerGlyValaHis 100  
 Db GGGCACTGGTATCCCAACCAAGATGGCTTAAGGGTTGGTCCAGATCCAGAGGCCCAT 362  
 QY 101 IleProLeuProProAlaProGlyGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
 Db ATTCCCTTACCAACAGCTCTGAGAGGCTGCCCTGGAGGCTGCTGCTCTGGGCAATTGGA 422  
 QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlyGlyHisProCysLeuVala 140  
 Db GGGCACTTTCCTCCACAGGGGCAAGCTCACTGCCGAGAGACCCGTGCTTAAAGGCT 482  
 QY 141 ProArgLeuArgCysGlnSerCysGlyLeuValaGlnAlaProGlyLeuArgAlaGlyGly 160  
 Db CCACGGCTAGATGCCAGTCTCTCAAGCTGGACAGGCCCCGGGCTCAAGGGGAGGAA 542  
 QY 161 ArgSerProGlyGlyLeuSerLeuGlyArgArgGlyArgAsnValaAsnThrAlaProGln 180  
 Db AGGTTCACCAAGAGAGTCCCTGGGTGGGCGGTGGAAAGAAATTAATACGCCCCCAG 602  
 QY 181 PheGlnProProSerTrpGlnAlaThrValProGlyAsnGlnProAlaGlyThrProVal 200  
 Db TTTCACCCCCCAAGCTACCAAGGCCACAGTCCGAGAACCAAGCCACAGGCCACCCCTGTT 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGlyGlyValaGlyValArgLeuGlyTrpThrMet 220  
 Db GCATCCCTTAAGGGCCATCAACCCGAGCGAGGTGAGGAGGAGTGAATCTGAGATACACAAAG 722  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db GATGCCCTCTTGAATAGCGGCTCCAAACAGTTCTTCTCCCTGGAACCAAGTCACTGGTGA 782  
 QY 241 ValThrThrAlaGlyGlyLeuAspArgGlyThrThrLysSerThrHisValPheArgValThr 260  
 Db GTAAACCAAGCGGAGGAGGAGTGTGATCGTAGAACCAAGACCAACCTCTTCAAGGGTCAAG 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db GCGCAGAGACCAAGCGATGCCCAAGAGAGTGGCTGTACACTCACTTGTGGTACT 902  
 QY 281 AspThrAsnAspHisAspProValPheGlyGlnGlnGlyTrpLysGlyLeuSerLeuArgGln 300  
 Db GACACCAATGACATGACCTGTGTTGAGAGGACAGAGATCAAGAGAGGCTCAGGGAG 962  
 QY 301 AsnLeuGlyValaGlyTrpGlyValLeuThrValaArgAlaThrAspGlyAspAlaProPro 320  
 Db AACCTGGAGGTGGCTATAGAGTGTCACTGTCAAGGCCACGATGTGATGCCCTCCCTCC 1022  
 QY 321 AsnAlaAsnIleLeuTrpArgLeuLeuGlyGlySerGlyLysSerProSerGlyValaPhe 340  
 Db AATGCCAATATTTCTGACCGCTCTGAGAGGGGTCTGGGGGAGACCCCTCTGAAGCTTT 1082  
 QY 341 GlnIleAspProArgSerGlyValaIleArgThrArgGlyProValAspArgGlyGlyVala 360  
 Db GAGATCGACCTGTGCTCTGGGGGATCCGAACCCGTGGCCCTGTGATCGGGAGAGGGT 1142

QY 361 GluSerTrpGlnLeuThrValaGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db GAATCTTACCAAGTGAAGGTAGAGGCAAGTGAACAGGGTCGGGACCGGGGTCTCGAGAT 1202  
 QY 381 ThrThrAlaValaPheLeuSerValaGlnAspAsnAspAsnAlaProGlnPheSer 400  
 Db ACCACAGCGGTGTTTCTTCTCTGTGAGAGTGAACAAATGAATATGCCCCCAAGTTTACT 1262  
 QY 401 GlnLysArgTrpValaValaGlnValaArgGlyAspValaThrProGlyValaProValaLeuArg 420  
 Db GAGAGCGCTATGTGCTCCAGGTGAGAGATGTGATCTCCAGGGGCCCACTGATCCCA 1322  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValaHisIleTrpSerIleMetSer 440  
 Db GTCAACAGCTCGGATCGAGAACCAAGGGAGCAATGCGTGTGCTCACTATGACATGAGAT 1382  
 QY 441 GlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyValaLeuAspValaSer 460  
 Db GGCAATGCTCGGGGACAGATTATCTGATGTCAGATGCCAGACTGTGATGTGGTGAAC 1442  
 QY 461 ProLeuAspTrpGlyThrThrLysGlyTrpThrLeuArgValaArgAlaGlnAspGlyGly 480  
 Db CCTCTTGAATAGAGACACCAAGAGATGACCTTACGGGTGCAAGCACAGATGTGGC 1502  
 QY 481 ArgProProLeuSerAsnValaSerGlyLeuValaThrValaGlnValaLeuAspIleAsnAsp 500  
 Db CGTCCCCCACTCTTAATGTCTGTGGTGTGAGACATGAGTCTGGATATCAACAGAC 1562  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValaLeuGlySerValaProLeu 520  
 Db AATGCCCATCTTGTGTCAGACCCCTTTCAGGGTACTGTCTGAGAGGCTCCCTTA 1622  
 QY 521 GlyTrpLeuValaLeuHisValaGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db GGTACCTGTTCTCTCAATGTCAGGCTATCAAGCTGATGCTGTGACATATCCCGCTG 1682  
 QY 541 GlnTrpArgLeuAlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db GAATACCCCTTGTGCTGGGTGGGACATGATCTTCCATCAACATAGGACACAGGC 1742  
 QY 561 TrpIleSerValaAlaAlaGlyLeuAspArgGlyGlyValaAspPheTrpSerPheGlyVala 580  
 Db TGGATCTCTGTGGCTGTGTAATGAGCCGAGAGGAGTGAATTTCTACAGCTTGGGTA 1802  
 QY 581 GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValaSerValaThrVala 600  
 Db GAACTCGAGACCAATGGCACTCCAGCACTCACTGCTGGCAAGTCAAGGTGACTGTG 1862  
 QY 601 LeuAspValaAsnAspAsnAsnProThrPheThrGlnProGlyTrpThrValaArgLeuAsn 620  
 Db CTGGATGTCAACGACCAATCCAACTTTACCCCAACCAAGATACACAGTGGGCTCAAT 1922  
 QY 621 GlnAspAlaAlaValaGlyThrSerValaValaThrValaSerAlaValaAspArgAspAlaHis 640  
 Db GAGATGAGCTGTGGGACCAAGGTGTGACGGGTGACGTGTGAGACGTGATGCTCAT 1982  
 QY 641 SerValaIleThrTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db AGTGCATCACTTACCAATCAACAGTGGCAATCTGCAAAACCGTTCTCATCAACCAAGC 2042  
 QY 661 GlnSerGlyGlyGlyLeuValaSerLeuAlaLeuProLeuAspTrpLysLeuGlyGlnArg 680  
 Db CAAGTGTGTGTGGGTGTGATCTTCCCTGCGCACTGAGCTCAAACTTGAAGGGGAG 2102  
 QY 681 TyrValaLeuAlaValaThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValaVala 700  
 Db TATGTTGGCTGTTAACGCTCCGATGACCTGGCAGGACGAGACGACAAATGTGTG 2162  
 QY 701 AsnValaThrAspAlaAsnThrHisArgProValaPheGlnSerSerHisTrpThrValaAsn 720  
 Db AATGTCACGACGCAACACCACTGCTCTTTCAGAGCTCCCACTATACAGTGAAT 2222  
 QY 721 ValaGlnLeuAspArgProAlaGlyThrThrValaValaLeuIleSerAlaThrAspGlyLeu 740

Db 2223 GTTATATGAGACCGCGCGGAGGACCAACGCTGGTGGATCAGCGGCAACGAGAGAGAC 2282  
 Qy ThrGluValAspAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 Db 2283 ACAGGTGAGAAATGCCCGCATCACTTCAATGAGAGACAAATCCCGCATTCGCGATC 2342  
 Qy AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGlyAspGlnValSer 780  
 Db 2343 GATCGACAGACGCGGGGCTGTCAACCAACGAGCTGAGCTACGAAAGCAAGTGTCT 2402  
 Qy TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIlySerAspThrThrTyr 800  
 Db 2403 TACACCCCTGGCCATTATCTGCTCGGAGCAATGCAATCCCGAAGTCCGACCACTTAC 2462  
 Qy LeuGluIleLeuValAspAspValAspAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db 2463 CTGAGATCTCTGGTGAACGACGTGAATGAAATGCCCTCAGTTCCTGGCGACTCCCTAC 2522  
 Qy GlnGlySerValTyrGlyAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2523 CAGGGCAGTGTCTATGAGAGATGTGGCACTTCACTAGCGTCTGCAAGTCTCAGCACT 2582  
 Qy AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2583 GATGCTGATTCTGACCTTAATGAGGAGGTCTTCAACCTTCCAAAGGAGGCGACGATGGA 2642  
 Qy AspAlaAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgAspLeuAsp 880  
 Db 2643 GACCGTACTTATTTGTTAGTCCACGTCGAGGCACTGTCGCAAGCGCTACGAGAGCTGAT 2702  
 Qy ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro 900  
 Db 2703 CGAAGAAACGCGGCGCATATGTCTGGGCGCATATGCAATGAGCAAGGGAGTCCCCCA 2762  
 Qy AlaArgThrProMetGluValThrValThrValLeuAspValAspAspAsnProProVal 920  
 Db 2763 GCCCGCACCTAATGAAAGTACAGTCACTGTGTGATGATGAAATGCAATCCCCCTGTC 2822  
 Qy PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 Db 2823 TTTAGACAGATGATGATTGATGTGTGTGTGTGAAAGAAAGAACCCCATTTGGGCTAGCCGTG 2882  
 Qy AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 2883 GCCCGGTCACAGCCACTGACCCCGATGAAAGCAACATGCCAGATTATCAATTCACAAATT 2942  
 Qy ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980  
 Db 2943 GTGAGGGCAACATCCCTGAGGCTCTCCAGCTGACATCTTCTCGGGGAGCTGACAGCC 3002  
 Qy LeuValAspLeuAspTyrGlyAspAspProGluTyrValLeuValIleGlnAlaThrSer 1000  
 Db 3003 CTGGTAAAGCTTAAGCTACGAGGACCGGCTCAGTACGTCCTGCTGCTCATCCAGGCCACGTCA 3062  
 Qy AlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 3063 GCTCTCTGTGTAGCGCGGCTACAGTCCAGCTCCGCTCTTGGACCGCAATGACAAACCCA 3122  
 Qy ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040  
 Db 3123 CAGAGTGGGCAATCTTGAATCTTTTCAACAACATATGTCACCAATCGCTCAAGGACG 3182  
 Qy PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 3183 TTCCCTGGGGGTGCAATTGGCGAGTACCTGCCCATGACCCTGATCTCAGATAGCTG 3242  
 Qy ThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3243 ACTTACAGCTTTGAGCGGGGAAATGAATCAAGCTGTGCTGCTCAATGCTTCCACGAGGT 3302  
 Qy GluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100

Db 3303 GACGTGAGCTTAAGCCGCGCATGGAACAACACCGGCTCTGGAGGCAATCATGACGTG 3362  
 Qy LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
 Db 3363 CTGGTGTCAACCGGCTTACACAGCTGACCGGCCAAGTGGCGCTGGGTGAGCAATCATC 3422  
 Qy ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
 Db 3423 ACCGATGAGATGCTCACCAACAGCATCAGCTGCGCTGGAGGACATATGTCACCGAGCGC 3482  
 Qy PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
 Db 3483 TTCCTGTACCACTGTCTAAGCCCTTTCATCCAGGCGGTGGCCGACGCTGGCGACGCCA 3542  
 Qy ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
 Db 3543 CCGAGCACGTGTGTGCTTCAACGTAACGAGGAAACAGACCCCGCGGGGCGCAATC 3602  
 Qy LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu 1200  
 Db 3603 CTCAACGTAGCTGTGCGGTGGGCGAGCCGCGGAGCGCCGCGGGGCGGCCCTTCTCG 3662  
 Qy ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
 Db 3663 CCTCTGAGGACCTTGAGAGAGCGCTTATCAACCGAGCGCTGTGACGCGCATCTCG 3722  
 Qy AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240  
 Db 3723 GCACAGCGGTGTGCGCTTTCAGACGAAACATCTGCTCGGAGGCCCTGCGAAGACTAC 3782  
 Qy MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
 Db 3783 ATGCGCTGCGTGTGCGTGTGCTGCACTCCTCGCGGCCCTTCAATCCGCTCTCTCC 3842  
 Qy ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
 Db 3843 GTGCTCTTCGCGCCATCAACCCCGTGGAGGGCTGCGCTGCGCGCCGCGCGGCTTC 3902  
 Qy ThrGlyAspTyrCysGlnThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3903 ACGGTGACTATCTACGAGAACGAGAGTGACCTCTGTACTCGCGGCGCTGTGGCCCCCAG 3962  
 Qy GlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 Db 3963 GGGCGCTGCGCGCGCGGAGGGCGGCTACACCTGCTGTGTGATGAGCTACAGGGGT 4022  
 Qy GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIysAsnGly 1340  
 Db 4023 GAGCACTGTAGGTGAGTGTGCTGCTGAGGCGCTTGACCCCGGGGTGTGCAAGAAATGG 4082  
 Qy GlyThrCysValAsnLeuLeuValGlyGlyPheIlyCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACCTGTGTCAACTGCTGTGTGGCGGTTCATATGAGATTTGCCCATCTGGAGACTTC 4142  
 Qy GluIlyProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 Db 4143 GAGAAACCTTACCTGACGAGGTGACACGCGCAGCTTCCCGGCCCATCTTCATCAACCTTT 4202  
 Qy ArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAlaThrIlyGluArg 1400  
 Db 4203 CCGGCGCTGGCGAGCGTTCATCTTCACTTCACTTGAGCCCTCTGTTTCCCAAGAGAGCGC 4262  
 Qy AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyHisAspPheValAlaLeuGlu 1420  
 Db 4263 GACGGGTGCTGTGTGACATGGGCGTTTCAATGAGAAACATGACTTGTGGCCCTCGAG 4322  
 Qy ValIleGlnGluGluValGlnLeuThrPheSerAlaGlyGluSerThrThrValIleSer 1440  
 Db 4323 GTATTCAGAGAGAGGTCCAGCTCACTTCTGCGAGGGAAGTCAACACAGGAGTCC 4382  
 Qy ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuIlySerTyr 1460  
 Db 4383 CCAATTGTGCGCGAGAGAGTCAATGATGCGCAATGCGTGCATACGCTGCAAAATCTAC 4442

QY 1461 AsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnLysValAla 1480  
 DB 4443 AAAPAGCACTGTGGGTGACAGACGGGGCTCCACAGGGCCCATCAGAGCAGAAAGTGGCT 4502  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB 4503 GTGGTAACCTGGATGGCTGTGACACAGAGTGGCTTGGCTTGGATCTGTCTGGAGC 4562  
 QY 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysSerLeuAspLeuThr 1520  
 DB 4563 AACCTACTCCTGTGCTGGCCAGGGGACCAGGGTGGCAGAAAGTCTTGATCTGAGC 4622  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGlnSerPheProValArgMetArg 1540  
 DB 4623 GGGCCCCCTGTACTAGCGGGGGGTGGCTGACCTGGCCAGAGACTTCCAGTCCCAATGGCG 4682  
 QY 1541 GluPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 DB 4683 CAGTTGGTGGCTGGACATGCGAAACCTGACAGTGAACAGCCGCACTAGACATGGCTGAC 4742  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
 DB 4743 TTCATTGGCAACATGGCACTGGCTGGCTGGCTGGCCAAAGAAAGTGTGTGACAGC 4802  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAspAlaPheSerCysGluCys 1600  
 DB 4803 AACACTGTGCACATGGGGCACTTGGTGAACCAAGTGGACCGCTGTCAAGCTGCGAGTGC 4862  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
 DB 4863 CCCCTGGGCTTTGGGGGCAAGACTGGCGCCAGAAATGGCAATCCACAGCACTTCCG 4922  
 QY 1621 GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpIleu 1640  
 DB 4923 GGCAGCAGCTGGTGGCTGGCAATGGCTCTGGCTGCCATCTCCAACTCGTGTACCTC 4982  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 DB 4983 AGCCTATGTCTCGACGGCGCCAGGGCGAGCGTGTCTGTGACGCATCACAAGGGGG 5042  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGlyThr 1680  
 DB 5043 CGCAGACCATCATCCCTTACAGCTACAGAGAGGCCCATGTATGCTGACGTTGAGGGGCA 5102  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTrpHis 1700  
 DB 5103 GGGCTTCAGGGCTCTCTCTCGTCTGGAGCCAGGCCGAGCCCAATGACGGTGACTGGCAC 5162  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 DB 5163 CATGCACACTGGCACTGGAGAGCAGCGGGGGCTGGCCATGCCATTCGTCTTCGAT 5222  
 QY 1721 TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 DB 5223 TATGGCAGCAGAGAGCAGAGGGCAACTGGGGCCCCCGGCTGATGTCTGACCTGAGC 5282  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaAlaArgLysPheArgGly 1760  
 DB 5283 AACATAACAGTGGGGCGGAATACCTGGGCCAGCGGGGTGGCCGTGGCTTGGGGC 5342  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer 1780  
 DB 5343 TGTTTGACAGGGTGTGGGTGAGCAGATACGCCAGAGGGGGTTAAACGCTGTGATCCAC 5402  
 QY 1781 HisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB 5403 CATGGGAGAGCATCACTGGAGCAAGGCTGTAGCTGCTGACCTTGTATCTCAAC 5462  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp 1820  
 DB 5463 CGGTGTCTGTAAAGCTATGTGACCAAGACTGGGAGCAGCTATTCTGTGAGCTGTGAT 5522

QY 1821 ProGlyTyrTrpGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnGln 1840  
 DB 5523 CAGGTATCATATGGTGAACAATGTATATGTGTGACTGAACCCGTTGTAGACCAAG 5582  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 DB 5583 TCTGTGTATCCCGCAAGGCCAGGTGCCCCCATGTGTATATCTGCAAGTGTCCCCCAAT 5642  
 QY 1861 TyrLeuGlyProTyrCysGlnThrArgLysAspGlnProCysProArgGlyTrpTrpGly 1880  
 DB 5643 TACCTGGGCACTTACTGTATGACAGATTTGACAGACTTGTCTCCGTGGGTGGAG 5702  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn 1900  
 DB 5703 CATCCACATGTGGCCCATGTGACATGTGATGTCAAGAAAGGCTTTGACCCAGCTGCAAC 5762  
 QY 1901 LysThrSerGlyGlyCysHisGlyCysLysGlnAsnHisGlyArgProGlnLysProThr 1920  
 DB 5763 AAGACAAAGCGGAGTGGCCATGTGCAAGAGAAACCACTACCGGCCCCAGGAGCCCAAC 5822  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940  
 DB 5823 TGCCTTGTGTGATGCTGTACCCCAAGGCTCTTGTCCAGAGTCTGTACCTTGAAGAT 5882  
 QY 1941 GlyGlnCysProCysLysAspProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 DB 5883 GGCAGTGTCCATGCAAGCAGAGTGTCACTGGGCGCTCAGTGTGACCGCTGTACAACCT 5942  
 QY 1961 PheAlaGlnValThrThrAsnGlyCysGlnLysValAsnTyrAspSerCysProArgAlaIle 1980  
 DB 5943 TTGGCTGAGGTACCAACCAATGGCTGTGAAGTAATTAAGACGTGCCCAAGCAAGCACT 6002  
 QY 1981 GluAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
 DB 6003 GAGGCTGGGATGTGGTGGCCCGGTACCCGCTTGGGGTGGCTGTCTGTCTCTGTCTC 6062  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnHisAspGlyTyrLeuProPro 2020  
 DB 6063 AAAGCTCTTTGGAGCATGTGTGGCGCATGTGTATGAGCAGAGGGGGTGGCTCCGCCA 6122  
 QY 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLysGlyPheAlaGlnArgLeu 2040  
 DB 6123 AACCTCTTCACTGACGTCACTCACTTCTCAAGATGAAGGGCTTGTGTAGGGCTA 6182  
 QY 2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg 2060  
 DB 6183 CAGCGGAATGAGTCAAGGCTGTAGCTCAAGGGGCTCCAGAGCTAGCCCTGTCTGTGCC 6242  
 QY 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
 DB 6243 AACGCCAGCAGCACAAGCTGGCTAATTCCGAGCGACGTCAAGAGTGGCTTACCAAGCTG 6302  
 QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 DB 6303 GCCAGCGGCTGTGGCCCAAGAGACCCAGCGGGCTTGTGGCTGTCTGCAACAG 6362  
 QY 2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
 DB 6363 GACGTGCACTTCACTGAAATCTGTGGGTGGGCAAGCGCTCTCTGAGACAGCCCAAC 6422  
 QY 2121 LysArgHisTrpGlnLeuIleGlnGlnThrGlnGlyGlyThrAlaTrpLeuLeuGlnHis 2140  
 DB 6423 AAGCGGCACTGGAGCTGATATCAGAGAGCAGAGGGTGGCACCGCTGTCTGTCCAGCAC 6482  
 QY 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisTrpTyrLeuSerProPhe 2160  
 DB 6483 TATGAGGCTTACGCCAGTGGCTGTGGCCCAAGAACATGCGGCACTTACTTAAGCCCTTC 6542  
 QY 2161 ThrIleValThrProAsnIleValIleSerValArgLeuAspLysGlyAsnPheAla 2180  
 DB 6543 AACATCGTCAAGCCCAACATTTGATCTCGTATGGGCTTGGACAAAGGAACTTGTCT 6602  
 QY 2181 GlyAlaLysLeuProArgTyrGlyAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200

```

Db 6603 GGGGCCAAGCTGCCCCCTACAGAGCCCTGCGGGAGAGCCGCCGAGCTTGAACA 6662
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220
Db 6663 ACGATCTTCTGCTGAGCTGTCTCTTCAAGAGAGCCGCCCTGGTCAAGCCGAGGC 6722
Qy 2221 ProGlyValIleLeuGluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeu 2240
Db 6723 CCGGAGAGAGCCCAAGAGCCAGAGAGAGCTGGCAAGGCAAGCAAGCAAGCCGAGAGCTG 6782
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleThrArgThrIleAlaGlyLeuLeuPro 2260
Db 6783 AGCCAGAGGTAGGCTGTGGCCAGGCTCATCTACAGCAAGCCGCGGCTACAGCTG 6842
Qy 2261 HisAsnTrpAspProAspLeuArgSerSerLeuArgValProIleAspProIleIleAsnThr 2280
Db 6843 CATACCTATACCTTACAGAGCCAGCTTGAAGTCCCAAGCCGATCATACAGCA 6902
Qy 2281 ProValValSerIleSerValHisAspAspGluGluLeuLeuProAlaLeuAspGly 2300
Db 6903 CCGGTGGTGAAGTCAAGCTGCTCATGATGAGAGCTTCTGCCCCGAGCCCTGAGCAAA 6962
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleProIleCysVal 2320
Db 6963 CCGGTACGGTGAAGTCCCGCTGCTGAGACAGAGAGGAGCAAGCCATCTGTCTC 7022
Qy 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyValTrpSerAlaArgGlyCysGlu 2340
Db 7023 TTCTGGAAACATTCACCTGCTGTCAAGTGCACAGTGGCTGTGTGCGGAGAGCTGAA 7082
Qy 2341 ValValPheArgArgGlnSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360
Db 7083 GTGCTCTCCGCAATGAGACCAAGCTCAGCTGCCAGTCAACCAATGAGACTTCGCT 7142
Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuArgThrLeuThr 2380
Db 7143 GTGCTACAGAGCTGTCTTCGCGGAGAAATGGAGAGTCTGCTCACTGAAGACATGACA 7202
Qy 2381 TyrValAlaLeuGlyValIleThrLeuAlaAlaLeuLeuThrPhePheLeuThrLeu 2400
Db 7203 TACGTGCTCTAAGTGTCACTTGGCTGCTGCTTGTCTCACTTCTTCTTCTCACTCTC 7262
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgLeuLeuThrAlaAlaLeuGly 2420
Db 7263 TTGGGTATCTCGGCTCAACCAACAGGCACTCCGATACCTGACAGCTGCGCTGAGC 7322
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db 7323 CTGGCTCAGGCTGGCTTCTCTCTCTGAGAAATCAACAGGCTGACCTCTTGTGCTGACA 7382
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuThrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460
Db 7383 GTGATTCACATCCGCTGCACTTCTGACCTCTGCACTTTCCTGCGCTGTGAGAG 7442
Qy 2461 AlaLeuHisLeuThrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480
Db 7443 GCGCTTGACCTGATCCGAGCACTCACTGAGGTGCGCATGTCAACACCGGCCCATGCGC 7502
Qy 2481 PheTrpTrpMetLeuGlyTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db 7503 TTCTACTACATGCTGGGCTGGGCGGTGGCTGCTTCAATCAAGGAGCGAGGCGCTG 7562
Qy 2501 AspProGluGlyTrpGlyAsnProAspPheCysTrpLeuSerIleIleTrpAspThrLeuIle 2520
Db 7563 GACCCCAAGGCTACGAGGAACCTGACTTGTGCTGCTCTCATCTATGACAGCTCATC 7622
Qy 2521 ThrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrpIleLeu 2540
Db 7623 TGGAGTTTGTGGCCCGGTGCTTGTGCTGTCAAGTGTCTTCCGTGATCTCTG 7682
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluValGlyProValSer 2560

```

```

Db 7683 GCGGCCCGGAGCTCTCTGTGCTGCCAGCGGCAAGGCTTTGAGAGAAAGTCTGTCTCG 7742
Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580
Db 7743 GGCCTGCAAGCTCTTGTGGCGGTCTCTGTGTGAGGCGCACGTGGCTGTGCACTG 7802
Qy 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTrpLeuPheAlaThrCysAsnCysIle 2600
Db 7803 CTCTGTGTCACAGAGAACCTCTCTTCCACTACTCTTGTGCTACTGCAATGATGATC 7862
Qy 2601 GlnGlyProPheIlePheLeuSerTrpValValLeuSerGlyValArgValAlaLeu 2620
Db 7863 CAGGCGCCCTTCAATCTTCTCTCTATGTGTGTGTACAGAGAGTCCGGAACACTTC 7922
Qy 2621 LysLeuAlaCysSerArgLysSerProAspProAlaLeuThrThrLysSerThrLeu 2640
Db 7923 AAGCTTGCTGAGCGCCAGCCAGCCCTGACCTGTCTGACCAAGTCCACCTG 7982
Qy 2641 ThrSerSerTrpAsnCysProSerProTrpAlaAspGlyArgLeuTrpGlnProTrpGly 2660
Db 7983 ACCTGCTCTACACTGCCCCAGCCCTACGAGATGGGCGGCTGTACAGCCCTACGGA 8042
Qy 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTrpIle 2680
Db 8043 GACTGCGCGGCTCTCTGACAGCACAGTGTGCTGGGCGAGAGTCAAGCCAGCTACATC 8102
Qy 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700
Db 8103 CCTCTTGTGAGGAGAGAGTCCGCACTGAACTTGGCCAAAGGCCCTCGCTGAGG 8162
Qy 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnHisAspProAspThrAspSer 2720
Db 8163 GATCCAGGAGGCTGTCTTCTGAGAGTCAAGCACAGCAATGATCTGACACAGCACTCC 8222
Qy 2721 AspSerAspLeuSerLeuGluGluAspAspGlnSerGlySerTrpAlaSerThrHisSerSer 2740
Db 8223 GACAGTGAACCTGTCTTGAAGAACACAGAGTGTCTCTATAGCTCTTACCACTCATCA 8282
Qy 2741 AspSerGluGluGluGluGluGluGluGluGluGluAlaAlaPheProGlyValGluGlnGly 2760
Db 8283 GACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8342
Qy 2761 TrpAspSerLeuLeuGlyProGlyValAlaGluArgLeuProLeuHisSerThrProLysAsp 2780
Db 8343 TGGATAGGCTGTCTGTGGCTGAGACAGAGACAGTCCCTGCAAGTACTCCCAAGAT 8402
Qy 2781 GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800
Db 8403 GGGGCCCCAGGAGCTGTGGAGAGCCCTGTGCGAGAGACTTTGGAGCACAGCAAAAGAG 8462
Qy 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820
Db 8463 AGTATGAGGACAGGGGCCCCCTGAGAGCGGCTGCGGAGAAATGAGAGATGCCCTGTCTGA 8522
Qy 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisAlaGlyIleLeuLys 2840
Db 8523 GAGGGGTCTTGGGCCCCCTTCAAGCTCTTGTGCCAGCCCTCAAAAGGCAATCTTAAAG 8582
Qy 2841 LysLysCysLeuProThrIleSerGlnLysSerSerLeuLeuArgLeuProLeuGlnGln 2860
Db 8583 AAGAGTGTCTGCCACATCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642
Qy 2861 CysThrGlySerSerArgLysSerSerAlaSerGluGlySerArgGlyValProProPro 2880
Db 8643 TGCACAGGCTTCTTCCCGGAGGCTCTCCGCTATGAGGAGAGCGGAGGCGGCCCCCTCC 8702
Qy 2881 ArgProProProArgGluSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2900
Db 8703 CGCCCAAGCCCGGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8762
Qy 2901 SerIleLysAlaGlyThrValAspGluAspSerSerGlySerGluPheLeuPheAsn 2920
Db 8763 AGCATCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8822

```

QY 2921 PheLeuHis 2923  
 Db 8823 TTCCTGCAT 8831

## RESULT 3

AS11677  
 ID AS11677 standard, cDNA, 8871 BP.

AC AS11677;

DT 24-OCT-2001 (first entry)

DE Human Flamingo cDNA sequence.

XX Flamingo, human, splice variant; G-protein coupled receptor; diabetes;  
 KW signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
 KW obesity; hypotension; hypertension; urinary retention; angina pectoris;  
 KW myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
 KW benign prostatic hyperplasia; psychotic disorder; neurological disorder;  
 KW manic depression; delirium; dementia; severe mental retardation; ss;  
 KW Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KW antifungal; antiviral; antiparasitic; anti-HIV; anorectic; antianorectic;  
 KW antiparkinsonian; carbanit; cerebroprotective; neuroprotective;  
 KW antidepressant; anticonvulsant; antisense therapy; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..8771

FT /tag= a /product= "Human Flamingo protein #1"

XX WO200161003-A1.

XX 23-ANG-2001.

XX 19-FEB-2001; 2001WO-GB00680.

XX 19-FEB-2000; 2000GB-0004196.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Testa TT;

XX WPI; 2001-502792/55.

XX P-SDB; AAU07053.

XX An isolated Flamingo polypeptide useful for treating diseases such as

XX HIV, cancer, asthma, Parkinson's disease, acute heart failure,

XX osteoporosis -

XX Claim 5; Page 25-28; 66pp; English.

XX The sequence represents a cDNA which encodes a human Flamingo  
 CC polypeptide. Flamingo is a member of the G-protein coupled receptor  
 CC family, which is involved in signal transduction pathways. By screening  
 CC to identify compounds that stimulate or inhibit the function or level of  
 CC the protein, treatments can be developed for various diseases and  
 CC bacterial, fungal, protozoan and viral infections, including HIV, cancer,  
 CC diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease, acute  
 CC heart failure, hypotension, hypertension, urinary retention, angina  
 CC pectoris, myocardial infarction, stroke, ulcers, allergies and benign  
 CC prostatic hyperplasia. Also treatable are psychotic and neurological  
 CC disorders such as anxiety, schizophrenia, manic depression, delirium,  
 CC dementia, severe mental retardation, Huntington's disease and Gilles de  
 CC la Tourette's syndrome.

XX Sequence 8871 BP; 1712 A; 2859 C; 2587 G; 1713 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 8871

Score: 15518.50 Matches: 2923  
 Percent Similarity: 98.88% Conservative: 0  
 Best Local Similarity: 98.88% Mismatches: 0  
 Query Match: 99.83% Indels: 33  
 DB: 22 Gaps: 1

US-09-916-849a-3 (1-2923) x AS11677 (1-8871)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 1 ATGCGAGGCCCGCCACCGGCGCTCCCTCCCAAGCGCCGCCGCTGCTGCTG 60  
 QY 21 LeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 QY 41 GlySerArgGlyValGlySerSerGlyValCysAlaProMetGlyTyrLeuCysProSer 60  
 Db 121 GGGTCCAGGAGACGAGGCTTTCGGGGGCTGCGCCCATGGGCTGCTGCTGCTGCT 180  
 QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyTyrGlyLeuThr 80  
 Db 181 TGAAGCTCAACTCTGCTGCTCAACAGCCGCTGACGGAGATCGGGCACTGAGCTG 240  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100  
 Db 241 GGCACACTGGTACCCCAACAGATGGCTGAGGGTTGGTTCAGAAATCCGAGGCCCAT 300  
 QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
 Db 301 ATTCCTCAACACAGCTCTGAAAGCTGCCCCCTGAGCTGCTGCTGCTGCTGCTG 360  
 QY 121 GlyHisLeuSerProGluGlyValSerLeuThrLeuProGluGluHisProCysLeuValAla 140  
 Db 361 GGCCACTTTTCCCAACAGGCGACGCTCACTGCCCGAGAGACCCGCTTAAAGCT 420  
 QY 141 ProArgLeuArgCysGlnSerCysAlaSerLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
 Db 421 CCACGGCTCAGATGCAAGTCTCGCAAGCTGGCAAGGCCCGCGGCTCGAGGCGAGAA 480  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgValAsnThrAlaProGln 180  
 Db 481 AGGTCAACGAAAGACTCCCTGGGTGGCGTCGAAAGGAAATGTAAATACAGCCCCC 540  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyTyrProVal 200  
 Db 541 TTCCAGCCCCCAGCTACCAAGCCCAAGTCCGAGAAACAGCCAGGAGCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyArgLeuGluTyrThrMet 220  
 Db 601 GCATCCCTGAGGGCCATGCAACCCGACGAGGGTGAAGGTGCACTGAGATACCATG 660  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
 Db 661 GATGCCCTTTGATAGCGCTCCACCAAGTTCTTCTCCGAGACCACTGATGATGCA 720  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrThrLysSerThrHisValPheArgValThr 260  
 Db 721 GTAACCAACGCGAGGAGCTGATGTGAACCAAGACACCAAGCTTTCAGGGTCAAG 780  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrTyrLeuValThr 280  
 Db 781 GGGCAGGACCAACGGGATCCCGACGAAAGTCCCTGGCTACCACTCATCTTGTGTTACT 840  
 QY 281 AspThrAspAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGlu 300  
 Db 841 GACACCAATGACCATGACCTCTGTGTTCCAGACGACGAGGTACAGAGAGCTTCAGGAG 900  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGAGAGTTGGCTATGAGGTGCTCACTGACAGGCGACAGATGATGATGCCCTCC 960  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGluValPhe 340



D	961	AAATCCAAATATTCGTACCGCTCTGAGGGGCTGGGGGCAACCCCTCTGAAGTCTTT	1020
Q	341	GIUIleasPProAgsSerGIValIleargThraArgIProValAAsPaRGluVal	360
D	1021	GAGATCAACCTCGCTCTGGGGTATCCGAACCCGTGGCCCTGTGATCCGGAAAGAGTG	1080
Q	361	GIUsErYrgIleuThraValGIUaIleSerApsGInIYArAapProGIYProAgsSer	380
D	1081	GAATCCACCAAGCTGAGGAGTGAAGCAAGGAGTGAAGGAGTCCGGGCTCTGGAGT	1140
Q	381	ThrThraIaIaValPheLeuSerValGIUaAsPaAsPaAnaIaProGInPheSer	400
D	1141	ACCAACCCGCTGTTCCTTCCTGTGGAGGATGACATATATATATATATATATATAT	1200
Q	401	GIUIYArGIYValValGIUaIleValArgIUAAsPValThraProGIYAlaProValLeuArg	420
D	1201	GAGAGGCGCTATGTGTCTGAGTGAAGGAGATGATGATCCAGGGGCCCAAGTCTCGA	1260
Q	421	ValThraIaSerAAsPaAsPAGlySerAnaIaValIleIleIleSerIleMetSer	440
D	1261	GTCAACACCTCGATGACAGCAAGGAGCAATGCCGTGTGCACTATGATCATGAGT	1320
Q	441	GIYAsPaIaArgGIYGIInPheTYrLeuAsPaIaGIInThraValIleUaAsPaValSer	460
D	1321	GGCATGCTGGGGACAGTTTATCTGATGCCAGCTGAGCTGTGAGTGTGAGC	1380
Q	461	ProLeuAsPTrYGluThraThraIleGIYTYrThraIleValArgIaGIaAsPGLY	480
D	1381	CTCTTACATATGAGACGACCAAGAGATACCTTACGGGTGACGACAGAGATGTGCGC	1440
Q	481	ArgProProLeuSerAAsPaIleSerGIYleuValThraValGIUaIleUaAsP	500
D	1441	CGTCCCACTCTTAATGTCTGTGGCTGTGTGACAGTACAGTCTCGATATCAAGAC	1500
Q	501	AsnaIaProIlePheValSerThraProPheGIaIaThraValLeuGIUaSerValProLeu	520
D	1501	AATCCCCCATCTGTGACAGCAACCCCTTCCAGGCTACTGTCTGAGAGAGGTCCCTTA	1560
Q	521	GIYTYrLeuValIleUaIleValGIaIaIleAsPaIaAsPaIaGIYAsPaAnaIaArgLeu	540
D	1561	GGCTACCTGGTTTCCATGTCCAGGCTATGACGCTGATGTGTGCAATGCCGCTG	1620
Q	541	GIUITYrArgLeuIaAGIYValGIYHleAsPheProPheThraIleAsnaGIYThGIY	560
D	1621	GAATACCGCTGTGTGGGTGGACATGATCTCCCTTCCATCAACATATGGACAGGC	1680
Q	561	TrpIleSerValIaIaIaGIUaIleUaAsPaRGIUGIUAAsPheTYrSerPheGIYVal	580
D	1681	TGATCTCTGTGGCTGTGACTGACCGGAGAGAGTGAATTCATCAAGCTTGGGGTA	1740
Q	581	GIUIaIaArgAAsPiAGIYThraProIaIleUaThraIaSerIaIaSerValThraVal	600
D	1741	GAAGCTCGAGACATGACACTCCAGCACTCACTCGCTCGGCAAGTCAAGGTGACTGTC	1800
Q	601	LeuAsPaValaAsnaAsnaAsnaProThraPheThraIleProGIYTYrThraValArgLeuAsn	620
D	1801	CTGATGTCAACGACCAACATCCAACTTATCCCAACAGAGTACAGAGTCCGCTCAAT	1860
Q	621	GIUAsPaIaIaValGIYThraSerValThraValSerIaIaValaIaAsPaAsPaIaIle	640
D	1861	GAGATGCACTGTGGGACCAAGGTGTGACGGTGTGACGTGTGACCGTGTGACT	1920
Q	641	SeValIleThraTYrGIaIleThraSerGIYAsnaThraArgAnaRgPheSerIleThraSer	660
D	1921	AGTGTATCACTACCAAGATCAACAGTGGCAATATCGAAACCGCTTCTCATCACAGC	1980
Q	661	GIUsErGIYGIYleuValSerIleUaIleUaProLeuAsPTrYrIleUaGIUaArgIn	680
D	1981	CAAGGTGTGTGGGTGTGTATCCCTTGTCCGTGACATGACAACTTGAACGGGAG	2040
Q	681	TYrValleuIaValaIleThraIaSerApsGIYThraRGIaAsPThraIleValIaVal	700
D	2041	TATGTGTGGCTGTATCCGCTCCAGTGGCACTCGGACAGACAGGACACATGTGTG	2100
Q	701	AsnaValThraAsPaIaAsnaThraIleArgProValPheGIaIleSerSerIleTYrThraIaAsn	720
D	2101	AATGTACCAAGCCCAACACCAATGCTCTGTCTTCAAGCTCCCATCTACAGAT	2160
Q	721	ValaAsnaAsPaArgProIaAGIYThraValIleUaIleSerIaThraAsPGIUaAsP	740
D	2161	GTTAATGAGAACCGGCGGACAGGACCAAGTGTGTGATGACGCCCAAGTGAAGAGC	2220
Q	741	ThraGIUaAsnaIaArgIleThraTYrPheMetGIUaSerSerIleProGIaIlePheArgIle	760
D	2221	ACAGGTGAATGTCCCGCATCACTTATCTTCAAGAGACAGATCCCAAGTCCGATC	2280
Q	761	AsPaIaAsPThraGIYAlaValThraThraIaGIUaIleUaAsPTrYGluAsPInaIleSer	780
D	2281	GATGACAGACCGGGGCTGTCAACCCAGGCTGAGCTGATGATGACAAAGTGTCT	2340
Q	781	TYrThraIleUaIleThraIaArgAsnaGIYIleProGIaIleSerAAsPThraTYr	800
D	2341	TACACCTGGCCATTACTGTCTGGGACATGAGCATTCACAGAGTCCGACACCTAC	2400
Q	801	LeuGIUaIleUaValaAsnaAsPaValaAsnaAsPaAnaIaProGIaIlePheUaArgApsEYr	820
D	2401	CTGAGATCTGTGTGACAGAGTGAATGACAAATGCCCTCAAGTCTGTGAGATCTTAC	2460
Q	821	GIaGIYSerValTYrGIUaAsPaValProPheThraSerValLeuGIaIleSerIaThra	840
D	2461	CAGGACAGTGTATGAGATGTGCCACCTTCACTAGGCTGTGAGATTCAGCACT	2520
Q	841	AsPaArgApsSerGIYleuAsnaGIYArgValPheTYrThraPheGIaIleGIYAsPaAsPGLY	860
D	2521	GATCGATYTCGTGACTTATGTGACAGGAGTCTTCAACCTTCCAAAGAGCCACATGGA	2580
Q	861	AsPGLYAsPheIleValGIUaIleUaSerThraSerGIYIleValArgThraIleUaArgLeuAsP	880
D	2581	GAGGTGACTTATTTATGTATGATCCAGTCAAGCATGTGTGCAAGCTTACGAGGCTGAT	2640
Q	881	ArgGIUaAsnaValaGIaIleTYrValleUaArgIaTYrAlaValaAsPGLYMetProPro	900
D	2641	CGAGAAAGTGTGGCCAGATGTCTTGGGGATATGACGTGACAAAGGGATGCCCA	2700
Q	901	AlaArgThraProMetGIUaIleThraValThraValIleUaAsPaValaAsnaAsnaProProVal	920
D	2701	GCCGCAACCTTATGTGAATGTGACATGTGTGTGAGATGTGAATATATCCCTGTGC	2760
Q	921	PheGIUaIleUaAsPGLYUaAsPaValPheValGIUaIleUaAsnaSerProIleGIYleuIaVal	940
D	2761	TTTGACCAAGATGATGTATGTGTGTGTGAAAGAGAACAGCCCATTTGGGCTAGCGTG	2820
Q	941	AlaArgValThraIaThraAsPProApsGIUGIYThraAsnaIaGIaIleMetTYrGIaIle	960
D	2821	GCCGGGTCAAGCCACTGACCCGATGAAAGGACCAATGCCAGTATGTATCCAGAT	2880
Q	961	ValGIUGIYAsnaIleProGIaIleUaIlePheGIaIleUaAsPGLYUaAsnaIaIleUaThraIa	980
D	2881	GTTGAGAGGACATCCCTTAAGTCTTCAAGCTGGAATCTTCCGAGAGTGAAGAGC	2940
Q	981	LeuValaAsPLeuAsPTrYGluAsPaRGIUGIYTYrValIleUaValIleGIaIaIleThraSer	1000
D	2941	CTGTGATGACTTGAACAGAGAGACGGGCTGAGTATGCTGTGTATCCAGGCAAGTCA	3000
Q	1001	AlaProLeuValaSerArgAlaThraValIleValArgLeuUaAsPaRgAsnaAsnaPro	1020
D	3001	GCTTCTGTGTGGCGGCTTCAAGTCAAGTCCGCTCTTCAACGCAATGACCA	3060
Q	1021	ProValleuGIYAsnaPheGIUaIleUaPheAsnaAsnaTYrValThraAsnaRgSerSer	1040
D	3061	CAAGTGTGGCACTTGTAGATCTTTCACACATATGTCAACATGTGCTCAAGCAGC	3120
Q	1041	PheProGIYGIYAlaIleGIYArgValProIaIleAsPProAsPTrIleSerApsSerLeu	1060
D	3121	TTCCCTGGGGGTGCATTTGGCGAGTACTGCCCATGACCTGTATCTCAATAGTCTG	3180



D	5341	CATGGGAGAGCAACAGTGGACCAAGCTGTAGCTGCTGACCTTGTGACTCAAC	5400
Q	1801	ProCyProAlaAaSerTYrCySerAsnApTTPaPseTYrSerCySerCyAsp	1820
D	5401	CCGCTCTCTGTAACAGCTATTGGACGACACCTGGACAGCTATTCTCTGACCTTGAT	5460
Q	1821	ProGlyTYrTYrGlyAspAsnCySerAsnValCyAspLeuAsnProCyGlyLysGln	1840
D	5461	CAAGGTACTAGTGTGACAACTGACTAATGTGTGACCTGACCACTGGTAGACACAG	5520
Q	1841	SerValCySerThrArgLysProSerAlaProHisGlyTYrThrCySerGluCyProProAsn	1860
D	5521	TCGTGTGTACCCGACAGCCGACGAGCCGCCATGCTATCTGCGAGTGTCCCCCAAT	5580
Q	1861	TYrLeuGlyProTYrCySerGluThrArgGlyLeuAspGlnProCyProAspGlyTTPTrGly	1880
D	5581	TACCTTGGGCACTACTGTGACACAGATTTACAGACTTGTCCCCGTGGTGGGGA	5640
Q	1881	HisProThrCySerGlyProCyAsnCyAspValSerLysGlyPheAspProAspCyAsn	1900
D	5641	CATCCACATGTGGCCCATGTGACCTGTGATGTGACAAAGGCTTTGACCCAGACTGCAC	5700
Q	1901	LysThrSerGlyGluCySerHisCySerLysGluAsnHisGlyTYrArgProProGlySerProThr	1920
D	5701	AAGACAGCGCGAGTCCACTGACAGAGAACCACTACCGGCCCCAGGACAGCCCCACC	5760
Q	1921	CysLeuLeuCyAspCySerTYrProThrGlySerLeuSerArgValCyAspProGlyLysP	1940
D	5761	TGCTCTTGTGTACTGTGACCTGACCCACAGGCTCTTGTTCAGAGTCTGTGACCTGAGAT	5820
Q	1941	GlyGlnCySerProCySerLysProGlyValLysGlyArgGlnCyAspArgCyAspAsnPro	1960
D	5821	GGCAGAGTGCATGACCAAGCAGGTGTGATCGGGGTGATGACCGCTGTGACCAACCT	5880
Q	1961	PheAlaGluValThrThrAsnGlyCySerGluValAsnTYrAspSerCyProArgAlaIle	1980
D	5881	TTTGCTAAGGTCAACCAACAAAGCTGTGAAGTGAATTATGACACTGCCACAGACGATT	5940
Q	1981	GluAlaGlyLysLeuTYrProArgTYrArgPheGlyLeuProAlaAlaIleProCyPro	2000
D	5941	GAGGCTGGGATCTGTGGCCCCCGTACCGCTTGGGGTGTGTGCTCTCCCTGCTCC	6000
Q	2001	LysGlySerPheGlyTYrAlaValArgHisCyAspGlnHisArgGlyTYrPLeuProPro	2020
D	6001	AAAGGCTCTTTGGGACGTGCTGTGGCCACTGTGATGACACAGGGGTGGCTCCCCCA	6060
Q	2021	AsnLeuPheAsnCySerThrSerIleThrPheSerGlyLeuLysGlyPheAlaGluArgLeu	2040
D	6061	AACTCTTCACTGACGTCATCACTTCTCAGAACTGAAGGCTGTGGTAGGGGCTA	6120
Q	2041	GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg	2060
D	6121	CAGGGATGAGTGAAGGCTTGACTCAGGGGCGCTCCAGAGCAGTACGCTGCTCGCG	6180
Q	2061	AsnAlaThrGlnHisGlnAlaGlyTYrPheGlySerAspValLysValAlaTYrGlnLeu	2080
D	6181	AAACCCACGACGACACAGCTGCTACTTGTGGCAGCAGTCAAGGTGGCTTCCACAGCTG	6240
Q	2081	AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLysSerAlaThrGln	2100
D	6241	GCCACGGGCGCTGGCCCAAGAGACCCAGCGGGGCTTTGGGCTGTCTGCCACACAG	6300
Q	2101	AspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
D	6301	GACGTGACACTTCACTGAGAACTCTGCTGGGTGGGACGCGCTCTGACACAGCCAAC	6360
Q	2121	LysArgHisGlyProGlyLeuLysGlnGlnThrGlnGlyGlyTYrAlaAlaPLeuLeuGlnHis	2140
D	6361	AAGGGGACCTGGGAGGTGATTCAGACACAGAGGGGTGGCACCGCTGGCTCCACAC	6420
Q	2141	TYrGluAlaTYrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTYrLeuSerProPhe	2160
D	6421	TATGAGGCTTACGCCAGTGGCTTGGCCCAAGCATGCGGACACCTTAAAGCCCTTC	6480
Q	2161	ThrIleValThrProAsnIleValIleSerValValArgLeuAspLysGlyAsnPheAla	2180
D	6481	ACCATGTGACCGCCCAACATTTGTCACTTCCGTATGTCGCTTGGACAAAGGAACTTGTCT	6540
Q	2181	GlyValAlaLysLeuProArgTYrGlyAlaLeuArgGlyGlyGlnProProAspLeuGlyThr	2200
D	6541	GGGGCCAAAGCTGCCCCGCTTACAGAGGCGCTGCTGGGGAGAACACCCCGGACCTTGAGACA	6600
Q	2201	ThrValIleLeuProGlySerValPheArgGlyThrProProValValArgProAlaGly	2220
D	6601	ACAGTATTTCTGTGAGTGTGTCTTCAAGAGAGAGCCCGGTGTCAAGGCCGAGGC	6660
Q	2221	ProGlyGlyAlaGlnGlyProGlyGlnLeuAlaArgArgGlnArgArgHisAspProGlyLeu	2240
D	6661	CCCGAGAGAGCCAGAGAGGACAGAGAGCTGGCACCGGCAAGACGACCCGAGGAGCTG	6720
Q	2241	SerGlnGlyGluAlaValAlaSerValIleIleTYrArgThrLeuAlaGlyLeuLeuPro	2260
D	6721	AGCCAGGGTAGAGCTGTGGCCAGCGTCACTACCTACCGCCTGGCCGCTACTGCTCT	6780
Q	2261	HisAsnTYrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr	2280
D	6781	CATACTATGACCTTGACACAGCGCACTTGAGAGTCCCAAGCCCGATCATCAACACA	6840
Q	2281	ProValIleSerIleSerValHisAspAspGlnGlyLeuLeuProArgAlaLeuAspLys	2300
D	6841	CCGCTGTAGGATCAGCTGCTCATGATGATGAGAGCTTCTGCCCCGGGCTCGACAA	6900
Q	2301	ProValThrValGlnPheArgLeuLeuGlyThrGlnGlyArgTYrLysProIleCySerVal	2320
D	6901	CCCGTCAACGCTGACGTTCCGCTGTGAGACAGAGAGACGACCAAGCCATCTGTGTGC	6960
Q	2321	PheTTPAsnHisSerIleLeuValSerGlyTYrThrGlyTYrTrpSerAlaArgGlyCySerGln	2340
D	6961	TTCTGACCACTTCAATCTGTGCTAGTGGACAGAGTGGCTGTGTGGCCAGAGGCTGTGA	7020
Q	2341	ValIleValPheArgAsnGlySerHisValSerCySerGlnCyAsnHisMetThrSerPheAla	2360
D	7021	GTCGTCTTCGCAATGAGAGCCACGTCAGCTCCAGTGGACCACTGACAGAGCTTCGCT	7080
Q	2361	ValLeuMetAspValSerArgArgGlyLysGlnGlyLysLeuProLeuLysThrLeuThr	2380
D	7081	GTGCTCATGAGAGCTTCTGCGGGAGAGATGGAGATCTCTCCACTGAAGACACTGACA	7140
Q	2381	TYrValAlaLeuGlyValThrLeuAlaIleLeuLeuLeuThrPhePheLeuThrLeu	2400
D	7141	TACGTGGCTCTGAGTCACTTGGCTGCTCTTGTCTCACTTCTTCTTCTCACTCTC	7200
Q	2401	LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleLeuGly	2420
D	7201	TTGCGTATCTCGCTGCCCAACCAACCGGCACTCCAGCTGACACTGACCTCCCTGGGC	7260
Q	2421	LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCySerThr	2440
D	7261	CTGGCTCAGCTGCTCTTCTCTGGAGATCAACAGGCTGACCTCCCTTTTGTCTGACA	7320
Q	2441	ValIleAlaIleLeuLeuHisPheLeuTYrLeuCySerThrPheSerTrpAlaLeuLeuGln	2460
D	7321	GTCATTGGCATCTGTGTGACTTCCGTATCCTGTGACCTTTTCCGGGCTGTGCGAG	7380
Q	2461	AlaLeuHisLeuTYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
D	7381	GCCCTTCACTTATCCGCGGACCTCACTGAGGTGGGAGTCAACACCGGCCCCAGCCGC	7440
Q	2481	PheTYrMetLeuGlyTYrTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
D	7441	TTTCTACTACATGTGGGCTGGGGCGTGGCTGCTTCATATCAAGGGCTTACCGTGGGCTG	7500
Q	2501	AspProGlyGlyTYrGlyAsnProAspPheCySerTrpLeuSerIleTYrAspThrLeuIle	2520
D	7501	GACCCGAGGGGCTAGGGAAACCTGACTTCTGCTGCTCTCATATATGACAGCTCATC	7560



CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotoxic,  
 CC antianemic, antipneumatic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndrome, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
 CC hepatitis. AX3440-ABX3485 encode the MDR polypeptides represented in  
 CC AB011550-AB011845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 11762 BP; 2286 A; 3738 C; 3283 G; 2455 T; 0 other;

# Alignment Scores:

Pred. No.:	0	Length:	11762
Score:	15393.00	Matches:	2894
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	59.02%	Indels:	0
DB:	25	Gaps:	0

US-09-916-849A-3 (1-2923) x ABX34546 (1-11762)

QY 30 LeuGlyAspGluValGlyProCysArgSerLeuGlySerArgGlyValArgGlySerSerGly 49  
 DB 2 TTGGAGAGCCAGAGTGGGGCCCTGCTGCTTGGGGTCCAGGGAGCGAGGCTCTTGGGG 61  
 QY 50 AlaCysAlaProMetGlyTyrLeuGlyProSerSerAlaSerAsnLeuTyrLeuTyrThr 69  
 DB 62 GCTGCGCCGCCAAGGGCTGGCTCTGTCACTCTCAAGGCTGGAACCTTGGCTCTCAACC 121  
 QY 70 SerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuValProHisAspGly 89  
 DB 122 AGCCGCTGCAGAGTGGCGGCACTGAGCTGAGCTGAGCTGAGTACCCACACAGATGGC 181  
 QY 90 LeuArgValTyrCysProGluSerGluAlaHisIleProLeuProProAlaProGluGly 109  
 DB 182 CTGAGGATTGGTGTCCAGAAATCCAGAGCCCAATATCCCTTACCAACAGCTCTGAGAGC 241  
 QY 110 CysProTyrSerCysArgLeuLeuGlyTyrLeuGlyHisLeuSerProGluGlyLeu 129  
 DB 242 TGCCCTCGAGCTGCTGCTCTCTGGGCACTTGGAGGCCACTTTCCCAAGGGAGCTC 301  
 QY 130 ThrLeuProGluGluHisProCysLeuValAlaProArgLeuArgCysGlnSerCysLys 149  
 DB 302 ACACGCGCCGAGAGCACCCGCTTAAAGGCTCCACCGCTCAGATGCCAGTCTCGCAG 361  
 QY 150 LeuAlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGly 169  
 DB 362 CTGGACAGGCCCCGGGCTCAGGGCAGGGGAAAGTCAACAAAGATCTCTGGGATGG 421  
 QY 170 ArgArgGlySerArgAsnValAsnThrAlaProGlnPheGlnProProSerTyrGlnAlaThr 189  
 DB 422 CGTGGAAAGGAATGTAATACAGCCCCCAAGTCCAGGCCCCCAAGTCAACAGGCCA 481  
 QY 190 ValProGluAsnGlnProAlaGlyThrProValAlaSerLeuArgAlaIleAspProAsp 209  
 DB 482 GTGCGGAGAACCAACCCAGAGCAACCCCTGTGATCCCTGAGGGCCATGAGCCCGAC 541  
 QY 210 GluGlyGluAlaGlyValArgLeuGluTyrThrMetAspAlaLeuPheAspSerArgSerAsn 229  
 DB 542 GAGGCTGAGGCGAGCTGAGTGAATACCATGATGCTCTTTGATAGCCGCTCCAC 601  
 QY 230 GlnPheSerSerLeuAspProValThrGlyAlaValThrThrAlaGluGluLeuAspArg 249  
 DB 602 CAGTCTTCTCTCCCTGAGCAAGTCACTGATGAGTAAACACAGCCGAGAGTGGATGCT 661  
 QY 250 GlnThrLysSerThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArg 269

DB 662 GAGACCAAGAGCACCAACGCTTCAAGGTCAAGGGCAGAGACCAAGGCAATGCCCGACCA 721  
 QY 270 SerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAspAspHisAspProValPhe 289  
 DB 722 AGTGGCTTGGCTTACCTCACTCAACATCTTGTTATCGAACCAATACCAATGACCTTGTT 781  
 QY 290 GluGlnGlnGluTyrLysGluSerLeuArgGluAsnLeuGluValGlyTyrGluValLeu 309  
 DB 782 GACACGACGAGATCAAGAGAGCTCAGAGGAACTTGGAGGTTGGCTATGAGGTGCTC 841  
 QY 310 ThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeu 329  
 DB 842 ACTGACAGGGCCACGAGATGGTATGCTCCCTCCCAATGCCAAATATTCGTGATCCGCTG 901  
 QY 330 GluGlySerGlySerProSerGluValPheGluIleAspProArgSerGlyValIle 349  
 DB 902 GAGGGGTCTGGGGGAGCCCTCTGAAGTCTTGAATGACCTGCTCTGGGGATGATC 961  
 QY 350 ArgThrArgGlyProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAla 369  
 DB 962 CGAACCCGTGGCTCTGTGATGGGAAAGGTGGAATCTTACCAAGCTGACGTAGAGGCA 1021  
 QY 370 SerAspGlnGlyValArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerVal 389  
 DB 1022 AGTGACCAAGGTGGGACCCGGGCTCTCGAGTACCAAGCCGCTGTTTCTTCTGCTG 1081  
 QY 390 GluAspAspAspAspAspAspAlaProGlnPheSerGluLysArgThrValAlaGluValArg 409  
 DB 1082 GAGATGACATATATATATGCCCCAGTTTATGAGAAAGGCTATATGATTCAGATGAGG 1141  
 QY 410 GluAspValThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspLysGly 429  
 DB 1142 GAGGATGTGACTCCAGAGGGGCCAGTACTCCAGTCAAGCTCGAGTGAAGCAAGGGG 1201  
 QY 430 SerAsnAlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeu 449  
 DB 1202 AGCAATGCCCTGTGTGACTATATGATCATATGAGGAAATGCTGGGGAGAGTTTATCTG 1261  
 QY 450 AspAlaGlnThrGlyAlaLeuAspValIleSerProLeuAspTyrGluThrThrLysGlu 469  
 DB 1262 GATGCCAGCTGAGACTCTGATGTGTGAGCCCTTTCATATGAGACGACCAAGAG 1321  
 QY 470 TyrThrLeuArgValAlaArgAlaAspGlyValArgProProLeuSerAsnValSerGly 489  
 DB 1322 TACACCTTACGGGTGAGAGCAGATGGTGGCCGCCCTCCCACTCTTATATGCTCTGGC 1381  
 QY 490 LeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrPro 509  
 DB 1382 TTGGTGCAGGTACAGTCTCGATATCAACGAAATGCCCTTCCTTCGACACACCCCT 1441  
 QY 510 PheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAla 529  
 DB 1442 TTCCAGGCTACTGCTCTGAGAGCCGCTCCCTTAGCTACTGTTCTTCAATGTCCAGGCT 1501  
 QY 530 IleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlnHis 549  
 DB 1502 ATCGAGCTGATCTGTGTGACATGCCCGCTGAAATACCGCTCTGCGGGTGGACAT 1561  
 QY 550 AspPheProPheThrIleAsnAsnGlyThrGlyTyrIleSerValAlaIleGluLeuAsp 569  
 DB 1562 GACTTCCCTTCAACATCAACATGAGCAGGGCTGATCTCTGTGCTGTGAATCTGAC 1621  
 QY 570 ArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAla 589  
 DB 1622 CGGAGGAAATGATTTCTTACACCTTTGGGTGAGAGCTGAGACCATGSCATCCAGCA 1681  
 QY 590 LeuThrAlaSerAlaSerValSerValThrValLeuAspValAsnAspAsnProThr 609  
 DB 1682 CTCATGCTCTCGGCGAGTGTCACTGATCTGTGATTCACCAACCAATTCACACC 1741  
 QY 610 PheThrGlnProGluTyrThrValArgLeuAspGluAspAlaIleValGlyThrSerVal 629



Db 1742 TTtACCACCAAGATGACAGTGGGGCTCAATGAGATGACCTGTGGGCAACGCGT 1801  
Qy 630 ValThrValSerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSer 649  
Db 1802 GTGAGCGGTGTGAGTGTGAGCCGTGATGCTCATAGTGCATCACTCAACGATGTCAGAT 1861  
Qy 650 GtAAenThrArgAsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeu 669  
Db 1862 GGCATTAATCTCGAAACCGCTTCTCCATCAACAGCGAAAGTGTGTGGGTGTATCCCTT 1921  
Qy 670 AlaLeuProLeuAspTyrIleValLeuGluValArgGlnTyrValIleuAlaValThrAlaSerAsp 689  
Db 1922 GCCCTGCCACTGACCTCAAACTTAAAGCGGAGATGtGTGTGGCTGTATCCGCTCCGAT 1981  
Qy 690 GlyThrArgGlnAspThrAlaGlnIleValIAsnValThrAspAlaAsnThrHisArg 709  
Db 1982 GGCATCTGGGACGAGACCGGCAAGATTGTGTGAATGTCAACGAGCCCAACCCATGCT 2041  
Qy 710 ProValIlePheGlnSerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThr 729  
Db 2042 CCTGTCTTCAAGAGCTCCCACTATACAGTGAATGTTAATGAGAGACCGGCGGCAAGCAC 2101  
Qy 720 ThrValIleValIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyr 749  
Db 2102 ACGGTGTGTGATGATAGCGCCAGAGTGAAGACAGAGTGAATATGCCGCACTCAC 2161  
Qy 750 PheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThr 769  
Db 2162 TTCATGAGAGACAGATCCCGCATGTCGCGCATCGATGAGACAGGGGGCTGTCCACACC 2221  
Qy 770 GlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAsp 789  
Db 2222 CAGGTGAGCTGAGACTTACGAAACCAAGTGTCTTACCCCTGGCCATTACTGTGGGAC 2281  
Qy 790 AsnGlyIleProGlnIleSerAspThrThrTyrLeuGlnIleLeuValAsnAspValAsn 809  
Db 2282 AATGGCATTTCCCAAGATGCCACACCACTACCTGAGAGATCTGTGAACAGCTGAAT 2341  
Qy 810 AspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValPro 829  
Db 2342 GACAAATGCCCTCAGTTCCTGTGAGACTCTTACAGGGGAGTGTATGAGAGATGTGCA 2401  
Qy 830 ProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArg 849  
Db 2402 CCCTTCACTAGCTGTCTGACATTCACGACCACTGATCGTGAATTCGAGCTTAATGGCAGG 2461  
Qy 850 ValPheTyrThrPheGlnGlyGlyAspAspGlyAspPheIleValGluSerThr 869  
Db 2462 GTCTTCTACACTTCCAGAGAGGCGACAGTGAAGCGGTGACTTATGTGTGAGTCCACG 2521  
Qy 870 SerGlyIleValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeu 889  
Db 2522 TCAGGCAATCGTGGCAAGCTACGAGGCTGGATGAGAGAACTGTGCCAAGATGTCTTG 2581  
Qy 890 ArgAlaTyrAlaValAspGlyGlyMetProProAlaArgThrPrometGluValThrVal 909  
Db 2592 CGGGCATATGACGTGACAGAGGGAGTGCCTCCAGGCCGACCACTATGAAGTGCACGTC 2641  
Qy 910 ThrValIleAspValAsnAspAsnProProValIlePheGlnGluAsnGluPheAspValPhe 929  
Db 2642 ACTGTGTGTGATGATGAATGACATCCCTCTGTCTTGAAGAGAGATGTGATGTGTTT 2701  
Qy 930 ValGlnGluAsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAsp 949  
Db 2702 GTGGAAAGAGAACAGCCCATTTGGCTAGCGCGTGGCCGCGGTCAAGCCACTGACCCGAT 2761  
Qy 950 GlnGlyThrAsnAlaGlnIleMetTyrGlnIleValIleGlnGlyAsnIleProGluValPhe 969  
Db 2762 GAAGGACCAACAGTCCAGATTATGTACAGATTGTGGAGGGCMAATCCCTGAGGTCTTTC 2821  
Qy 970 GlnLeuAspIlePheSerGlyGlyLeuThrAlaLeuValAspLeuAspTyrGluAspArg 989  
Db 2822 CAGCTGACATCTTCTCCGGGGAGCTGACAGCCCTGTGATGACTTGACTTACAGAGAGCCGG 2881

Qy 990 ProGlnTyrValIleuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrVal 1009  
Db 2882 CCGTAGTACGTCTCTGTATCATCAGGCCACGTCACTCTCTGTGTAGCCGGCTACATCTC 2941  
Qy 1010 HisValArgLeuLeuAspArgAsnAspAsnProProValIleuGlyAsnPheGlnIleLeu 1029  
Db 2942 CAGCTCCGCTCTCTTGAACCGCAATGACAAACCAAGTGTGGGCAATTTGAGATCTCTT 3001  
Qy 1030 PheAsnAsnTyrValIleThrAsnArgSerSerPheProGlyGlyAlaIleGlyArgVal 1049  
Db 3002 TTCAACAATATGTACCAATGTCTCAAGCAGCTTCCCTGGGGGTGCTATGGCCGAGTA 3061  
Qy 1050 ProAlaHisAspProAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGlu 1069  
Db 3062 CCGGCCATGACCCGTGATATCTCAGATATGTCTGACTTACACTTTGAGCGGGAAATGAA 3121  
Qy 1070 LeuSerLeuValIleLeuAsnAlaSerThrGlyGluLeuIleuSerArgAlaLeuAsp 1089  
Db 3122 CTGAGCTGTGTCTGTCAATGTCCCTCAAGCGGTGAGCTGAAGCTTAAGCCGCGCACTGGAC 3181  
Qy 1090 AsnAsnArgProLeuGlnAlaIleMetSerValIleuValSerAspGlyValHisSerVal 1109  
Db 3182 AACAAACGGCTCTGAGAGCCCATATGACGTGTGTGTGACAGCGGTGACACAGCGTG 3241  
Qy 1110 ThrAlaGlnCysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIle 1129  
Db 3242 ACCGCCCAAGTGGCGGTGTGTGTGATCCATCATCAACGATGAGTCTCAACCAACGATC 3301  
Qy 1130 ThrLeuArgLeuGluAspMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPhe 1149  
Db 3302 ACGCTGGCGCTGAGGACATGTCAACCCAGAGGCTTCCGTGACCACTGAGCTGAGCTTTC 3361  
Qy 1150 IleGlnAlaValAlaIleThrLeuAlaThrProProAspHisValIleValPheAsnVal 1169  
Db 3362 ATCAAGCGGTGTGGCGGACAGCTGACAGCGCACCGGACCAAGTGTGTCTTAACGTA 3421  
Qy 1170 GlnArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGln 1189  
Db 3422 CAGCGGACACCGACGCGCCCGGGGGGCACATCTTCAACGTGAACCTGTGTGGTGGCAG 3481  
Qy 1190 ProProGlyProGlyGlyProProPheLeuProSerGluAspLeuGlnGluArgLeu 1209  
Db 3482 CCGCCAGAGGCGCGGGGGCGGGCGCTTCTGTCTCTGTAGAGACTGTGACAGAGCGCTTA 3541  
Qy 1210 TyrLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValIleuProPheAspAsp 1229  
Db 3542 TACTTCAACCGACCTGTGACGCGCATCTCGGACAGCGGTGTGCTTCTTGAAGAC 3601  
Qy 1230 AsnIleCysLeuArgGluProCysGluAsnTyrMetArgCysValSerValIleuArgPhe 1249  
Db 3602 AACATCTGCGTGGGGAGCCCTGTGAGAACTACATGCGGTGTGTGGTGTGGCTTC 3661  
Qy 1250 AspSerSerAlaProPheIleAlaSerSerSerValIleuPheArgProIleHisProVal 1269  
Db 3662 GACTCTTCGCGGCTTCAATGCGCTCTCTCTCTGTCTTCCGCGCCATCCACCCGCTC 3721  
Qy 1270 GlyGlyLeuArgCysArgCysProProGlyPheThrGlyAspTyrCysGlyIleuThrGluVal 1289  
Db 3722 GAGAGGCTGGGCTGCGCTGCGCGCCGCGCTTCAAGGTATCTACTCGAGACCGAGGCTG 3781  
Qy 1290 AspLeuCysTyrSerArgProCysGlyProHisGlyArgCysArgSerArgGluGly 1309  
Db 3782 GACCTGTGCTACTGCGCGCTGTGTGGCCCGCCAGCGGGGTGTGCGCACCGGAGGGGGG 3841  
Qy 1310 TyrThrCysLeuCysArgAspGlyTyrThrGlyGlyHisCysGlyValSerAlaArgSer 1329  
Db 3842 TACACTGTCTGTGTGTGATGTGTACACGGGTGAGCACTGTGAGTGAAGTGTCTGTCA 3901  
Qy 1330 GlyArgCysThrProGlyValCysIleAsnGlyGlyTyrCysValIleuLeuLeuValGly 1349  
Db 3902 GCGCTTGTACCCCGGGGTGTGTGAGAAATGGGGGCACTGTGTCAACTGTGTGTGGG 3961



QY	1350	GLYHELYSCYASAPCPSPROSGEGLYASPHGGLYUSPFOYTCYSGINVALTHTR	1369
DB	3962	GGTTTCAGAGTCGATGGATGGCCATCTCGAACACTTCGAGAGCCCTCTGCGAGGTACCAACG	4021
QY	1370	ARGSERPHEPROALAHISERPHELLETHRPHARGGLYLEUAXGGLNARGPHENISPH	1389
DB	4022	CGAGCTTCCCCGCCCACTCTTCAATCACTTTGGCGGCTCGGCCCAAGCTTTCACATTC	4081
QY	1390	THIRLEUALLEUSERPHEALETHRYSGIUDARGAPGILEULEULEUTRYASNGLYARG	1409
DB	4082	ACCTGGGCGCTCTGGTTGGCCACAAGAGGCGCACGGGTGCTGTTGTAAAGGCGGT	4141
QY	1410	PHENANGIULYSNHSAPPHVEVALALEUGLIUVALLIENGINGUNVALGLEUETHR	1429
DB	4142	TTCAATGAGAACATGACTTTGTGGCCCTCGAGGTGATCCAGAGACAGGTCACTCAACC	4201
QY	1430	PHESERLAGIYGLUSERTHTRTHRVASERPROPHEVALPROGIYGLYVALISERASP	1449
DB	4202	TTCTCTGAGGGAGGTCAACACCAACGGGTGCCCATTCGTGGCCGAGGAGTCAGTGAT	4261
QY	1450	GLYGLINTPRHSITPRVALGLEULEULYSRYTRYTHENLYSPROULEULEUGLINTHGLY	1469
DB	4262	GGCCAGTGGCATACGTGTCAGACTGAATAACTACATAGACCACTGTTGGGTCAACAGGG	4321
QY	1470	LEUPROGLINGIYPROSGEGLUGLILYVALALAVALLTHRVALASPGIYCYASPTH	1489
DB	4322	CTCCCAACGGGCCCATCAAGACAGAGGTGGCTGTGTGACCGTGAATGGCTGTGACACA	4381
QY	1490	GLYVALALALEUARGPHEGLYSEVALLEUGLYASNTYSERCYALALAGLNGLYTHR	1509
DB	4382	GGAGTGGCGTTCGCTTCGATCTGTCTCGGGCAACTACTCTGTGGTCCCAAGGCAACC	4441
QY	1510	GLINGLYSERLYSERYSESERLEUASPREUTHRGYPROULEULEUGLYVALPRO	1529
DB	4442	CAGGTGGCAGCAAGAACTCTTGAATTCAGCGGGCCCCCTGGTCACTHAGCGGGGTGCT	4501
QY	1530	ASPLEUPROGUSERPHEPROVALARGMETARGGLNPHVEVALGYCYMETARGASMLEU	1549
DB	4502	GACCTGGCCGAGACCTTCCAGTCCGATGCGGACGCGATTCGTGGCTGCATCCGAAACCTG	4561
QY	1550	GLINVALASPSERARGHISILEASPMERLISAPRHELEIALAANAMNGLYTHVALPRO	1569
DB	4562	CAGGTGACACGCCGACATGACATGGCTACTTATTCGCAACAATGGCAACCGTGCT	4621
QY	1570	GLYCYABPROALALYLYSASNVALLCYASPSERASNTHCYSHISAMNGLYLYTHR	1589
DB	4622	GGCTGCCCTTGCCAGAGAAACGTGTGTGACAGCAACTTGCCCAATGGGGGACATTGC	4681
QY	1590	VALASNGINTPRAPRALPHSESECYAGIUCYSPROLEUGLYPHEGLYGLYLYSESCYS	1609
DB	4682	GTCGAACAGTGGAGCGGTTAGCTGCAGAGGCCCTCGGGCTTTGGGGGGAAGACGTGC	4741
QY	1610	ALAGLNGIUMETALASNPROGLNHSIPHELEUGLYSESERLEUVALALATPRHSGLY	1629
DB	4742	GCCCAAGAAAGGCCCAATCCACACACATCTCTGGGACAGACCTGTGTGGCTGGCAATGC	4801
QY	1630	LEUSERLEUPOLLIESERGLNPROTPTLYLEUSERLEUWETPHEARGTHARGGLNALA	1649
DB	4802	CTCTGTGCGCCCATCTCCCAACCTCGTAACTCAAGCTCAAGTTCCGACCGCGCAAGGCC	4861
QY	1650	ASPGIYVALLLEULEUGNALALILETHXARGIYARGSERTRILITHRLEUENLEUARG	1669
DB	4862	GACGGTGTCTGTGACAGCCATCAACAGGGGGGCGACACACATCACTTCACAGTTACGA	4921
QY	1670	GLUGLYHSIVALMETLEUSERVALIGULYTHRGYLEUGNALIASERSELEUARGLEU	1689
DB	4922	GAGGGCCACGTGAATGCTAGCGGTGAGGGCACAGGGCTTCAGGGCTCTCTCTCCGCTCG	4981
QY	1690	GLUPROGLYARGALASNPSPGLYASPTPRHSISALISAGILEUVALALEUGLYLASER	1709
DB	4982	GAGCCAGGCCCGGCCAATGACGATGACTGGCAACATGCAAGCTGGCACTGGAGACAGCC	5041
QY	1710	GLYGLYPROGLYHISALAILLEUSERPHEASPTRYGLINGIINARGALIGULYASN	1729

Db	5042	GGGGGGCTGGCCACGATTCCTGTCTCTTGGATTATGAGCAGACGAGGAGGGCCAC	5101
QY	1730	LeuGIyProAlaGLeuHISglYLeuHISLeuSerAsn1LeThrValIGlYglYIleProGlyI	1749
Db	5102	CTGGGGCCCCGGCTGCACATGGCTGTCCACCTGACACACATTAACATGGGGCGGAAATCCGGG	5161
QY	1750	ProAlaGlyValAlaAlaArgYlPheArgGlyCYsLeuGlnGlyValAlaGValSerAsp	1769
Db	5162	CCACCCGGGGGTGGGGCCGTGGCTTTCGGGGGCTGTGTGACGGGTGTGGCGGGTAGAGGAT	5221
QY	1770	ThrProGlnGlyValAlaSerSerLeuAspProSerHISglYglYSerIleAsnValIGlnG	1789
Db	5222	ACGCCAAGGGGGGTTTAACAGCTTGAAATCCAGCCATGGGAGAGACATCACTGAGACAA	5281
QY	1790	GlyCYsSerLeuProAspProCYsAspSerAsnProCYsAspProAlaAsnSerTYrCYsSer	1809
Db	5282	GGCGTAGCCCTGCCTGACCCCTTGACTCAAAACCGGTGCTCGCTAACACTATTCGACG	5341
QY	1810	AsnAspTrpAspSerTYrSerCYsSerCYsAspProGlyTYrTYrGlyValAspAsnCYsThr	1829
Db	5342	AACCACTGGAGACACTATTCCTGCAGCTGGATCCAGATTACTATGGTAGACACTGTACT	5401
QY	1830	AsnValCYsAspLeuAsnProCYsGlnHISglYSerValCYsThrArgLYsProSerAla	1849
Db	5402	AATGGGTGACCTGAACCCGTGTGAGACACACAGCTGTGTGTACCCGCAAGCCAGTGC	5461
QY	1850	ProHISglYTYrThrCYsGlnCYsProProAsnTYrLeuGlyProTYrCYsGlnTYrArg	1869
Db	5462	CCCCATGGCTATACCTTGCGAGTGTCCCCCAATTACTTGGGCCATACGTGTGAGACAGG	5521
QY	1870	IleAspGlnProCYsProAlaArgLYrTrpTrpGlyVHISProThrCYsGlyProCYsAsnCYs	1889
Db	5522	ATTGACACAGCTTGTCCCGGTGGCTGTGGAGACATCCACATGTGGCCATGACACTGT	5581
QY	1890	AspValSerLYsglyPheAspProAspCYsAsnLYsThrSerGlyGlyVHISCYsLYs	1909
Db	5582	GATGTACAGCAAAAGGCTTTGACCCGACTGTCAACAAAGCAAGCGCGAGTCCACTGAG	5641
QY	1910	GluAsnHISLYrArgProProGlySerProThrCYsLeuLeuCYsAspCYsTYrProThr	1929
Db	5642	GAGAACCACTACCGGCCCCCAGAGGCCCACTCGCTCTTGTGTGACTGTACCCACA	5701
QY	1930	GlySerLeuSerArgValCYsAspProGlnAspGlyGlnCYsProCYsLYsProGlyVal	1949
Db	5702	GGCTCTCTGTGCACAGCTGTGACCCCTGAGAGATGGCAGTGTCCATCAAGCCAGGTGC	5761
QY	1950	IleGlyTrpGlnCYsAspArgCYsAspAsnProPheHISglYValIleThrAsnGlyCYs	1969
Db	5762	ATCGGGGCTCAGTGTGACCGCTGTGACCACTTTGCTGAGGTACACCAACCAATCGCTGT	5821
QY	1970	GluValAsnTYrAspSerCYsProArgAlaIleGluAlaGlyIleTrpTrpProArgThr	1989
Db	5822	GAAGTGAATTATGACACTGCGCCACGAGCGATGAGGCTGGGATCTGTGGCCCCGTAC	5881
QY	1990	ArgPheGlyLeuProAlaAlaAlaIleProCYsAspProLYsglySerPheGlyTYrAlaValArg	2009
Db	5882	CGCTTCGGGCTGCTGTGCTGTCTCCCTGTGCCAAAGGCTCTTTGGAGCTGCTGTGCG	5941
QY	2010	HISCYsAspGlnHISArgGlyTYrPLeuProProAsnLeuPheAsnCYsThrSerIleThr	2029
Db	5942	CACGTGTATGACACAGGGGGGTGCTCCCCCAAACTCTTCACTACATGACGTGATCAC	6001
QY	2030	PheSerGlyLeuLYsglyPheAlaGlnArgLeuGlnArgAsnGlyLeuAspSer	2049
Db	6002	TTCTCAGAACTGAAGGGGCTTCGCTGACCGGCTACAGGGGATATGATGAGGCTTAAGACTCA	6061
QY	2050	GlyTrpSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHISTrpAlaGlyTYr	2069
Db	6062	GGGGGCTCCACGACGTAAGCCCTGTCTCTGTGGACCCACGACGACACACAGCTGGCTAC	6121
QY	2070	PheGlySerAspValLYsValAlaTYrGlnLeuAlaThrArgLeuLeuAlaHISglYSer	2089

Dh 6122 TTCGGAGAGGAGCTCAAGGTGGCTTACAGCTGGCCACCGGGCTGCTGGCCACGAGACC 6181  
Qy 2090 ThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlyLysLeuLeu 2109  
Dh 6182 ACCCAAGCGGGCTTGGGCTGTCTGGCCACAGAGAGTGACCTTCACTGAGATCTGGCTG 6241  
Qy 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisPTrpGlyLeuLysIleGln 2129  
Dh 6242 CCGGTGGGAGCGGCTCTCTGGACACAGCCACAGCGGACCTGGAGAGCTGATCCAGCAG 6301  
Qy 2130 ThrGlnGlyGlyThrAlaTrpLeuLeuGlnHisPheGlyLysAlaThrAlaSerAlaLeuAla 2149  
Dh 6302 ACGAGAGGTGGAGCCCGCTGGCTCTCCAGCACTATGAGGCTTACGCGAGTGGCTGGCC 6361  
Qy 2150 GlnAsnMetArgHisPheThrLeuSerProPheThrIleValThrProAsnIleValIle 2169  
Dh 6362 CAGAACATGGGAGCACTTACCTTAAGCCCTTCAACCATGGTCAAGCCCAACATTGCATC 6421  
Qy 2170 SerValValArgLeuAspLysGlyAsnPheAlaGlyAlaLysLeuProArgTyrgLysAla 2189  
Dh 6422 TCCGATGTCGCTTGGACAAAGGAACTTGTGGGGCCAAAGCTGCCCGCTACGAGGCC 6481  
Qy 2190 LeuArgGlyGlyLysLysProAspLeuGlyLysThrValIleLeuProGlyLysSerValPhe 2209  
Dh 6482 CTGCGTGGGAGAGAGCCCGGACCTTGAAGACAAAGTCAATCTGTGCTTAAGTCTGCTTC 6541  
Qy 2210 ArgGlyLysProProValValArgProAlaGlyProGlyGlyLysAlaGlnLysProGlyLys 2229  
Dh 6542 ACGAGAGACGCCCCCGGTGGTCAAGCCCGGAGGCCCCGAGAGGCCCCAGAGAGCCAGAGAG 6601  
Qy 2230 LeuAlaArgArgGlnArgArgHisPTrpGlyLysSerGlnGlyLysAlaValAspVal 2249  
Dh 6602 CTGGCAGCGGAGCAGAGCAGGCAAGCCGAGCTGAGCAGGGTGAAGGCTGTGGCCAGCGTC 6661  
Qy 2250 IleIleTyArgThrIleuAlaGlyLysLeuProHisAsnTyArgProAspLysArgSer 2269  
Dh 6662 ATCATTTACCGAGCCCTGGGCTAGCTGCTCATACCTATGACCTTCAAGAGCCGACG 6721  
Qy 2270 LeuArgValProLysArgProIleIleAsnThrProValValSerIleSerValHisAsp 2289  
Dh 6722 TTGAAGATCCCAAAAGCCGATCATCAACCCGCTGTGAGCATCAGCGTCCATGAT 6781  
Qy 2290 AspGlyLysLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArgLeuLeu 2309  
Dh 6782 GATGAGAGACTTCTGGCCCGGGGCTGGCAAAACCGTCAAGTGTGCGGCTGGCTG 6841  
Qy 2310 GlnThrGlnGlyLysArgThrLysProIleCysValPheTrpAsnHisSerIleLeuValSer 2329  
Dh 6842 GAGACAGAGAGAGGAGCCAGGCCATCTGTCTTCTGGAAACCAATTCCTGGTCAGT 6901  
Qy 2330 GlyThrGlyGlyTrpSerAlaArgGlyCysGlyValValPheArgAsnLysSerHisVal 2349  
Dh 6902 GGCAAGAGGTGGCTGGTGGGCAAGAGCTGTGAAGTGTCTTCCGCAATGAGGCCACGTC 6961  
Qy 2350 SerCysGlnCysAsnHisPheThrSerPheAlaValLeuMetAspValSerArgArgLys 2369  
Dh 6962 AACTGCGAGTGAACCAATGACGAGCTTGGCTGTCTCATGAGAGGTTTCTGGGGGAG 7021  
Qy 2370 AsnGlyLysIleLeuProLeuLysThrLeuThrTrpValAlaLeuGlyValThrLeuAla 2389  
Dh 7022 AATGGGAGATCTCTGCACTGAAGACATGACATGAGGTCTTAAGTGTCACTTGGCT 7081  
Qy 2390 AlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnLys 2409  
Dh 7082 GCCCTTCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7141  
Qy 2410 GlyIleArgArgAsnLeuThrAlaAlaLeuGlyLysAlaGlnLeuValPheLeuLeuGly 2429  
Dh 7142 GGCATTCGAGCTTAACCTGACAGCTGCCCTGGGCTGGCTGGCTGGCTGGCTGGCTGGCT 7201  
Qy 2430 IleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeu 2449  
Dh 7202 ATCAACAGAGCTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7261

Qy 2450 TyrLeuCysThrPheSerTrpAlaLeuLeuGlyLysAlaLeuHisPheLeuThrArgAlaLeuThr 2469  
Dh 7262 TACCTCTGACCTTTTCTTGGGCTCTGGAGAGGCTTGGACCTTACCCGGGACATCACT 7321  
Qy 2470 GlnValAlaArgAspValAsnThrGlyProMetArgPheTyTrpMetLeuGlyTrpGlyVal 2489  
Dh 7322 GAGTGGCCGATGTCAACAGCGGCCCTTCACTACATCTGGGCTGGGGCGTG 7381  
Qy 2490 ProAlaPheIleThrGlyLysAlaValGlyLysAspProGlyLysIleGlyLysProAsp 2509  
Dh 7382 CTGCGCTTATACAGGGCTAGCCGTGGGCTGGAGCCCGAGGCTTACGGGAACTTGAGC 7441  
Qy 2510 PheCysTrpLeuSerIleTyArgThrLeuIleTrpSerPheAlaGlyProValAlaPhe 2529  
Dh 7442 TTCTGCTGGCTCTCATCTATGACAGCTCATCTGAGATTGTCGTGGCCGGTGGCTTT 7501  
Qy 2530 AlaValSerMetSerValPheLeuTyTrpLeuAlaAlaArgAlaSerCysAlaAlaGln 2549  
Dh 7502 GCCGTCTGATGAGTGTCTTCTGTACATCTGGGCGCCGGGCTCTGTGTGCTCCAG 7561  
Qy 2550 ArgGlnGlyPheGlyLysValGlyProValSerGlyLysGlnProSerPheAlaValLeu 2569  
Dh 7562 CCGCAGGCTTTGAGAAAGAGTCTGTCTGGGCTGGAGCCCTCTTGGCGCTGCTC 7621  
Qy 2570 LeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeu 2589  
Dh 7622 CTGCTGTGAGAGCCACCTGGCTGTGGCACTGCTCTGTTCACAGGACACCTCTTC 7681  
Qy 2590 PheHisTyLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTy 2609  
Dh 7682 TTCACATACCTCTTGTACCTGCAATTGATTCAGAGGCCCTTCACTCTCTCTCAT 7741  
Qy 2610 ValValLeuSerLysGlyValArgValAlaLeuLysLeuAlaCysSerArgLysProSer 2629  
Dh 7742 GTGGTCTTACAGAGAGGTCCGAAAGACTCAAGCTTGTCTGAGCGGAGAGCCAGC 7801  
Qy 2630 ProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyArgCysProSerPro 2649  
Dh 7802 CCGACCTGTGCTGACCAAGCAAGTCCACCTCTGACTCTGTCTTCAATGCCCCAGCC 7861  
Qy 2650 TyrAlaAspGlyAlaGlyLeuTyGlnProTyGlyAspSerAlaGlySerLeuHisSerThr 2669  
Dh 7862 TACGAGATGGGCGGCTGTATCACGCCCTTACCGAATCTGGCGGCTCTCTCAAGCAGC 7921  
Qy 2670 SerArgSerGlyLysSerGlnProSerTyTrpPheLeuLeuArgGlyLysSerAla 2689  
Dh 7922 AGTGGCTGGGAGAGAGTCAAGCCAGCTACATCCCTTCTTGTGAGGAGAGAGTCCGCA 7981  
Qy 2690 LeuAsnProGlyGlnGlyProProGlyLysAlaAspProGlySerLeuPheLeuGlyLys 2709  
Dh 7982 CTGAACCTGGCCAAAGGGCCCCCTGGCTGGGGGATCCAGGCAACCTCTTCTTGAAGT 8041  
Qy 2710 GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlyAspAsp 2729  
Dh 8042 CAAGACCAAGCAGATGATCTTCAACAGGACTCCGACAGTGACTGTCTTGAAGAAGCAG 8101  
Qy 2730 GlnSerGlySerTyTrpAlaSerThrHisSerSerAspSerGlyLysGlyLysGlyLys 2749  
Dh 8102 CAGAGTGGCTCTATAGCTTCAACCACTCATCAACAGTGAAGGAGAAAGAGAGAGAG 8161  
Qy 2750 GlnGlyLysAlaAlaPheProGlyGlyLysGlnGlyTrpAspSerLeuLeuGlyProGlyVal 2769  
Dh 8162 GAAAGAGAGGCGGCTTCCCTGGAGAGAGAGGCTGGATAGCTGCTGGGCTGGAGCA 8221  
Qy 2770 GlnArgLeuProLeuHisPheThrProLysAspGlyGlyProGlyLysValAspPro 2789  
Dh 8222 GAGAGACTGGCCCTGCAAGATCTCCCAAGGATGGGGGCTGGGAGAGAGAGAGAGAG 8281  
Qy 2790 TrpProGlyAspPheGlyThrThrAlaLysGlySerSerGlyAsnGlyValProGlyLys 2809  
Dh 8282 TGGCCAGAGACTTTGGGACCAAGCAAAAGAGTGTGTGGCAAGGAGGAGGAGAGAG 8341

QY 2810 ArgLeuArgGluAsnGlyAspAlaLeuSerArgGluGlySerLeuGlyProLeuProGly 2829  
 DB 8342 CGGCTCGGGAGATGAGATGCTCTCTCCAGAGGGGCTTACGCCCTTCCAGGC 8401  
 QY 2830 SerSerAlaGlnProHisIleGlyIleLeuLysLysCysLeuProThrIleSerGlu 2849  
 DB 8402 TCTTCGCCAGCTCACAAGGACCTTAAAGAAAGAGTCTGCTCCACATCAGCGAG 8461  
 QY 2850 LysSerSerLeuLeuArgLeuProLeuGluGlnCysThrGlySerSerArgLysSer 2869  
 DB 8462 AAGAGACGCTCTCTGCGCTCCCTCCGAGCAATGACAGGGTCTTCCGGGCGCTCTCC 8521  
 QY 2870 AlaSerGluGlySerArgGlyGlyProProProArgProProArgGlnSerLeuGln 2889  
 DB 8522 GCTAGTAGAGGAGCGCGGGCGGCCCTCCCTCCCGCCACCGCCCGGAGAGCTCCAG 8581  
 QY 2890 GluGlnLeuAsnGlyValMetProIleAlaMetSerIleLysAlaGlyThrValAspGlu 2909  
 DB 8582 GAGCAGCTGAAACGGGTGATGCCATCGCATGAGATCAAGGCAAGCGATGATGAG 8641  
 QY 2910 AspSerSerGlySerGluPheLeuPhePheAsnPheLeuHis 2923  
 DB 8642 GACTCTCAGGCTCCGAATTTCTTCTTACTTCTCTGAT 8683  
 RESULT 5  
 ABK15177  
 ID ABK15177 standard; DNA; 9321 BP.  
 AC ABK15177;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human REPT9 9 CDNA sequence.  
 XX  
 KW REPT9: human; antiinflammatory; cytosolic; immunosuppressive;  
 KW antiviral; anti-HIV; antiarthritic; anticonvulsant; nocotropic;  
 KW neuroprotective; antiallergic; antibody; immunogen; endometriosis;  
 KW gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KW Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KW endocrine disorder; hypochlamus disease; Kallman's disease;  
 KW autoimmune disease; inflammatory disease; infertility; receptor;  
 KW acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KW allergy; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KW systemic lupus erythematosus; cell proliferative disorder;  
 KW cancer; developmental disorder; Duchenne muscular dystrophy; gene;  
 KW Becker muscular dystrophy; neurological disorder; epilepsy;  
 KW Alzheimer's disease; Huntington's disease; reproductive disorder; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 233..9033  
 FT /tag= a  
 FT /product= "REPT9 protein"  
 XX  
 PN MO200198354-A2.  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001MO-US19942.  
 PR 21-JUN-2000; 2000US-214027P.  
 PR 25-AUG-2000; 2000US-228045P.  
 PR 12-DEC-2000; 2000US-255104P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Griffith JA, Kallik DA, Tribouley CM, Yue H, Nguyen DB, Tang YF,  
 PI Tai P, Policky JT, Azimzai Y, Lu DM, Graul R, Yao WG, Burford N,  
 PI Hafalia AJA, Baughn MR, Bandman O, Patterson C, Yang J, Xu Y,  
 PI Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Dugan BM, Lu Y,  
 XX WPI; 2002-090432/12.

DR P-PDB; AAU74826.  
 XX  
 PT Twelve human receptors (referred to as REPT9-1 to REPT9-12), useful in  
 PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
 PT proliferative (e.g. cancer) disorders -  
 XX  
 PS Claim 65; Page 151-154; 157pp; English.  
 XX  
 CC This invention relates to twelve human receptors CDNA sequences  
 CC referred to as REPT9-1 to REPT9-12), and the proteins encoded thereby.  
 CC The proteins of the invention may have antiinflammatory, cytosolic,  
 CC immunosuppressive, antiviral, anti-HIV, antiarthritic, muscular active  
 CC general, anticonvulsant, nocotropic, neuroprotective, antiallergic  
 CC activities. The sequences of the invention may be used to produce REPT9  
 CC agonists or antagonists, and the protein sequences may be used to raise  
 CC anti-REPT9 antibodies. These molecules and the REPT9 polynucleotides and  
 CC polypeptides of the invention are useful in the diagnosis, treatment and  
 CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,  
 CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),  
 CC endocrine (e.g. hypochlamus disorder, Kallman's disease), autoimmune/  
 CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS),  
 CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,  
 CC multiple sclerosis, systemic lupus erythematosus), cell proliferative  
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
 CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,  
 CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)  
 CC disorders. Numerous other examples of each disorder are given in the  
 CC specification. The present sequence represents the human REPT9 CDNA  
 CC sequence of the invention.  
 XX  
 SQ Sequence 9321 BP; 1780 A; 3008 C; 2776 G; 1757 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 0 Length: 9321  
 Score: 15279.00 Matches: 2893  
 Percent Similarity: 98.44% Conservative: 2  
 Best Local Similarity: 98.37% Mismatches: 14  
 Query Match: 98.29% Indels: 32  
 DB: Gaps: 8  
 US-09-916-849a-3 (1-2923) x ABK15177 (1-9321)  
 QY 1 MetArgSerProAlaThrGlyValProLeuProThr---ProProProLeuLeuLeu 19  
 DB 223 ATGCGAGACCGGCGCACCGGGGCTCCCTCCCAAGCGCGCGCGCTGCTGCTG 282  
 QY 20 LeuLeuLeuLeuLeuProProProLeuGlyAspGlnValGlyProCysArgSer 39  
 DB 283 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 342  
 QY 40 LeuGlySerArgGlyValGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysPro 59  
 DB 343 TTGGGGTCCAGGGAGACGAGGCTCTTCCGGGGGCTGCGCCCAATGGGCTGCTGCTCA 402  
 QY 60 SerSerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeu 79  
 DB 403 TCCTCAGCGTGAACCTCTGCTGCTCAACAGCGCTCAGAGATGCGGGCACTAGCTG 462  
 QY 80 ThrGlyHisLeuValProHisIleAspGlyLeuArgValTyrCysProGluSerGluAla 99  
 DB 463 ACTGGCACCTGGATCCCAACACATGCTGAGAGGTTGATGTCAGAAATCCAGAGCC 522  
 QY 100 HisIleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyTyr 119  
 DB 523 CATATTCCTCCACACACGAGCTCTGAAAGGCTGCCCTGAGCTGTCGCTCGGCAATT 582  
 QY 120 GlyGlyHisLeuSerSerProGluGlyLysLeuThrLeuProGluGlnHisProCysLeuLys 139  
 DB 583 GAGGAGCACCTTTCCTCCACAGGACGATCACTGCGCGAGAGACACCGCTTAAG 642  
 QY 140 AlaProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGly 159

Dh 643 GCTCCAGCGCTCAGATGCCAGTCTTCGACGCTGGACACAGGCCCGCGGCTCAGGCGACG 702  
Qy 160 GIUATSerProgluGluSerleuGlyValArgLysArgAsnValAsnThrAlaPro 179  
Dh 703 GAAAGTCAACGAAAGTCCCTGGGTGGGCGCTCGAAAGGAAATGTAATACACCCCC 762  
Qy 180 GluPheGlnProSerSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrPro 199  
Dh 763 CAGTTCAGCCCCCGCAGCTACCAAGCCGACAGTCCGGAGAACCAAGCCAGGACGCCCT 822  
Qy 200 ValAlaSerLeuArgAlaAlaAspProAspGluGluValArgLysLeuGluTyrThr 219  
Dh 823 GTTGATCCCTGAGGGCCATGACCCGACGAGGGTGAAGGCGATGAGTCACTGAGTACCC 882  
Qy 220 MetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGly 239  
Dh 883 ATGATGCCCTCTTGTATAGCCGCTCCACAGTTCCTTCTCCCTGAGCCAGTCACTGGT 942  
Qy 240 AlaValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgVal 259  
Dh 943 GCAGTACCAACAGCGAGAGCTGGATCGTGAACAAGACCAACAGTCTTCAGGGTTC 1002  
Qy 260 ThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuVal 279  
Dh 1003 ACGGCGACGACACAGGATGCCCGACGAGATGCCCTGGCTACCTACCATCTTGATT 1062  
Qy 280 ThrAspThrAsnAspHisAspProValPheGluGlnGlnLysTyrLysGluSerLeuArg 299  
Dh 1063 ACTGACCAACCATGACATGACCTCTGTGTCAGACGACGAGATGACAGAGAGGCTTCAGG 1122  
Qy 300 GluLeuLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaPro 319  
Dh 1123 GAGAACTCGAGGTGGCTATGAGGTGCTCACTGTCAGGGCCACGAGATGGATGCCCTT 1182  
Qy 320 ProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGluVal 339  
Dh 1183 CCAATGCAATATCTGTACCGCTGCTGAGGGGTCTGGGGGACGCCCTCTGAAATCTC 1242  
Qy 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
Dh 1243 TTTGAGATCGAACCTCGCTCGGGGTGATCCGAACCCCTGGCCCTGTGATCGGGAAGG 1302  
Qy 360 ValGluSerTyrGlnLeuThrValGluLaserAspGlnGlyArgAspProGlyProArg 379  
Dh 1303 GTGGATCTCAACAGCTACGCTAGAGGCAAGTGAACAGGGTCGGGACCCGGGCTCTCG 1362  
Qy 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAsnAlaProGlnPhe 399  
Dh 1363 AGTACCAACAGCGCGCTGTTCTCTTCTGTGAGGATGACAAATGATATGCCCCCAGTTT 1422  
Qy 400 SerGluLysArgTyrValValGlnValArgGluAspValThrProGlyValaProValLeu 419  
Dh 1423 AGTGAAGAGCCCTATGTGTCTCAGGTGAGGAGATGTGACTCCAGGGGCCCAAGTACTC 1482  
Qy 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValHisTyrSerIleMet 439  
Dh 1483 CGAGTCAACAGCTCGATCGACGACAGGGGACAAATGCCGTGGTGCATTAAGCATCATG 1542  
Qy 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
Dh 1543 AGTGGCAATGCTCGGGGACAGTATTATCTGATGCCAGACTGGAGCTCTGATGTGTG 1602  
Qy 460 SerProLeuAspTyrGluThrThrLysGluTyrThrLeuArgValArgAlaGlnAspGly 479  
Dh 1603 AGCCCTCTTGAATATGACACACAGAGTACACCTTACGGGTGCCAGCACAGATGGT 1662  
Qy 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499  
Dh 1663 GGCGGTCCTCCCACTCTAATGTCTCTGGCTGGGACAGTACAGTCTCTGATATCAAC 1722  
Qy 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
Dh 1723 GACAAATGCCCAATCTTGTGTGACAGCCCTTTCAGGCTACTGTCTGGAAGGCTCCCC 1782

Qy 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
Dh 1783 TTAAGCTACCTGGTCTCTCATGTCTCAGGCTATGCAAGCTGATGGTGGACAAATGCCCG 1842  
Qy 540 LeuGluTyrArgLeuValArgValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
Dh 1843 CTGGAATACCGCTTGTCTGGGGTGGACATGACTTCCCTTCAACATCAACAAATGGGACA 1902  
Qy 560 GlyTyrPheSerValAlaAlaGluLeuAspArgGluValaAspPheTyrSerPheGly 579  
Dh 1903 GGCTGATCTCTGGGTGCTGAACTGACCGGAGGAAATGATTTCTACGCTTGCGG 1962  
Qy 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
Dh 1963 GTAGAGCTGAGACCAATGGCATCTCCAGCACTACCTCGGACAGTGTACAGCTGATCT 2022  
Qy 600 ValLeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeu 619  
Dh 2023 GTCTTGATGTCAACGACAAATCCAACTTATCCCAACAGATGACAGTGGGCTC 2082  
Qy 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValaAspArgAspAla 639  
Dh 2083 AATGAGATGCAAGCTGTGGCACCAAGCTGTGACGCTGTGACCTGTGAGCCGTGATGT 2142  
Qy 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
Dh 2143 CATGTGTATCACTACCAATCAATCCAGTACCAAGTGGCAATCTGAAACCGCTTCCATCAC 2202  
Qy 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArg 679  
Dh 2203 AGCCAAATGGTGGTGGGTATATCCCTTGCCCTGCACTGAGCTACAAACTTGAGCCG 2262  
Qy 680 GlnTyrValLeuValAlaThrAlaSerAspGlyThrArgLysAspThrAlaGlnIleVal 699  
Dh 2263 CAGTATGTGTGCTGTATCCGCTCCGATGGACCTCGGACGACACAGGACAGATTTGTG 2322  
Qy 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
Dh 2223 GTGAATTCACCAACGACCAACCCATCGTCTGTCTTCAAGACTCCCATATACAGTG 2382  
Qy 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
Dh 2383 AATGTATATGAGACCGGCGGACGACGACGAGTGTGATCAACGCGCACAGATGAG 2442  
Qy 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
Dh 2443 GACACAGGTGAGATGCCGCACTCACTTCAATGAGGACAGCATCCCCAGTTCCGC 2502  
Qy 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
Dh 2503 ATGATGACACACCGGGGCTGTACCAACCCAGCTGAGCTGAGCTACAGAGACCAAGTG 2562  
Qy 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThr 799  
Dh 2563 TCTTACACCTTGCCCAATATGCTCGGGCAATGGGATTCCTCCAGAAATCGACACACC 2622  
Qy 800 TyrLeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuAspSer 819  
Dh 2623 TACCTGAGATCCTGTGGAGACGATGAATGAATGAAATGCCCTCAATTCGCAAGACTCC 2682  
Qy 820 TyrGlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
Dh 2683 TACCAAGGAGGTCTATAGATGTGCCACCTTCACTAGCTCTCGACATCTCAGCC 2742  
Qy 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyLysAspAsp 859  
Dh 2743 ACTGATGTGATTTCTGAGACTTATGCAAGGCTCTTCAACCTTCCAAAGAGCGACGAT 2802  
Qy 860 GlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeu 879  
Dh 2803 GAGACGAGTACTTATATGTGAGTCAAGTCAAGGACATGTGCAAGCTTACGAGGCTG 2862

QY 880 AspArgGluAsnValAlaGlnIleuValIleuArgAlaIleuValAlaAspIleuSerPro 899  
Db 2863 GATCGAGAGAGACCTGGCCAGTATGCTTCCGGGCAATATGCAATGAGCAAGGGAGGCC 2922  
QY 900 ProAlaArgThrProMetGluValIleuValIleuAspValIleuAspAsnProPro 919  
Db 2923 CCGAGCCCGGACACCTATGGAAGTGAACAGTCACTGTGTGATGTGATGATGCAATCCCTC 2982  
QY 920 ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleuIleuAla 939  
Db 2983 GTCTTTGAGCAGAGATGATGATTTGTTGTGGAAGAGAACGCCCTTGGCTAGGCC 3042  
QY 940 ValAlaArgValIleuIleuAspProAspGluGluIleuAsnAlaGlnIleuMetIleuVal 959  
Db 3043 GTGGCCCGGGTCAAGCCACTGACCCCGATGAAGGCAACATCCCGAATTAATGACCAAG 3102  
QY 960 IleValGluGluValAsnIleProGluValPheGluIleuAspIlePheSerGluIleuThr 979  
Db 3103 ATGTGAGAGGGCAACCTCCCTGAGGTCTTCACCTGACATCTTCCGGGGAGCTGACA 3162  
QY 980 AlaIleuValAspIleuAspIleuValAspArgProGluIleuValIleuValIleuAlaIleu 999  
Db 3163 GCCCTGTAGACTTAACTAGAGACCGGCTGAGTACCTGTGATCAAGGCCAAG 3222  
QY 1000 SerAlaProIleuValSerArgAlaIleuValIleuValArgIleuAspArgAsnAspAsn 1019  
Db 3223 TCAAGCTCTCTGTGAGCCGGGCTACAGTCACTCCGCTCTTGAACCGCAATGACCAAC 3282  
QY 1020 ProProValIleuGluValAsnPheGluIleuPheAsnAsnIleuValIleuAsnArgSer 1039  
Db 3283 CACCAAGTGTGGGCAACTTGTGATCTTTCACCAACATATGCAACATCCCTCAACG 3342  
QY 1040 SerPheProGluValAlaIleuValArgValProAlaIleuAspProAspIleuSerAspSer 1059  
Db 3343 AGCTTCCCTGGGGGTGCAATGGCCAGATCACTGCCATGACCTGATATCTCAAGTAAAT 3402  
QY 1060 LeuThrIleuSerPheGluArgIleuValIleuSerIleuValIleuAsnAlaSerThr 1079  
Db 3403 CTGACTTACAGCTTGTGAGCGGGAAATGAATCACTGAGCTGCTCAATGCTCCACAG 3462  
QY 1080 GlyGluIleuIleuSerArgAlaIleuAspAsnAspArgProIleuValIleuMetSer 1099  
Db 3463 GGTGAGCTGAAGCTTAAGCCCGGCACTGAGCAACACCGGCTGTGAGGCAATCAAGAGC 3522  
QY 1100 ValIleuValSerAspGluValIleuSerValIleuValAlaIleuValIleuValIleu 1119  
Db 3523 GTGCTGTGTCAAGACCGGCTACACAGGCTGACCGCCAGTCCGCTGCTGACATC 3582  
QY 1120 IleThrAspGluMetIleuThrIleuSerIleuIleuArgIleuValIleuAspSerProGlu 1139  
Db 3583 ATCAACCATGATGCTCAACCAAGCATCACTGCTGCTGAGGACATGTCAACCCAG 3642  
QY 1140 ArgPheIleuSerProIleuGluIleuPheIleuAlaIleuAlaIleuAlaIleuAlaIleu 1159  
Db 3643 GGGTCTCTGACACACTGCTAAGCTTCACTCAAGGCGGTGGCGGCAAGCTGGCCAGC 3702  
QY 1160 ProProAspIleuValIleuAsnValIleuAsnValIleuAspThrAspAlaProGluValIleu 1179  
Db 3703 CCAACCGGACACCGGTGTCTTCAAGTACAGGGGACACCAACCGCCCGGGGGCCAGC 3762  
QY 1180 IleIleuAsnValSerIleuSerValIleuValProProGluValIleuValProProPhe 1199  
Db 3763 ATCTCAACATGAGCTGTGTGTGGGCAACCGGCAAGGGCCGGGGGGCGGCTTTC 3822  
QY 1200 LeuProSerGluAspIleuGlnIleuArgIleuIleuAsnArgSerIleuIleuThrAlaIleu 1219  
Db 3823 CTGGCCCTTAAGACCTGCAAGAGCGCTTAATCTCAACCGCAAGCTGTCTGAGGGCAATC 3882  
QY 1220 SerAlaGlnArgValIleuProPheAspAspAsnIleuValIleuArgIleuProGluValIleu 1239  
Db 3883 TCGGCAACAGGCGGTGCTGCTTCAAGCAACATCTGCGGGAGGCGCTCGAGAGAC 3942  
QY 1240 TyrMetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleuAlaSerSer 1259

Db 3943 TACATGGCGGTGTGCTGGGTGCTGGCTTCACTCTCCGGGCTTATGAGCTCTCC 4002  
QY 1260 SerValIleuPheArgProIleuIleuProValGluValIleuValArgCysArgProProGlu 1279  
Db 4003 TCCGTGCTCTTCCGGCCATCAACCCCGTGAAGGCTGCGCTGCGGCGCCCGGCGC 4062  
QY 1280 PheThrGluAspIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1299  
Db 4063 TTCAAGGTGACTTGTGAGACCGAGGTGAGCTCTCTCACTCCGCGCTCTGAGGCC 4122  
QY 1300 HisGluArgCysArgSerArgGluValIleuValIleuValIleuValIleuValIleuValIleu 1319  
Db 4123 CACGGGCGCTGCGGACCGGAGCGGCTTCAACCTGCTCTGCGTGAATGCTACAG 4182  
QY 1320 GlyGluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1339  
Db 4183 GGTGAGCACTGTGAGGTGAGTCTGCTCAAGCCCTTCAACCCCGGTGTCTGCAAGAT 4242  
QY 1340 GlyGluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1359  
Db 4243 GGGGCACTGTGTCAACTGTGTGGGCGGCTTCAAGTGGATGCTCCATCTGAGAC 4302  
QY 1360 PheGluValProValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1379  
Db 4303 TTGAGAGAGCTTACTGCAAGGTGACACCGGCACTCCCGCCACTTCAATCAC 4362  
QY 1380 PheArgGluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1399  
Db 4363 TTTCCGGGCTGTGGCAAGCTTCACTTCACTCCGCTCTCTTGTGCAAAAGAG 4422  
QY 1400 ArgAspGluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1419  
Db 4423 CCGGACGGGTGTCTGTATCAATGGGCTTCAATGAGAGCATGCTTGTGGCCCTC 4482  
QY 1420 GluValIleuGlnIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1439  
Db 4483 GAGGTGATCCAGAGCAAGTCACTCACTTGTGAGGGAGTCAACCAACGAGTG 4542  
QY 1440 SerProPheValProGluValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1459  
Db 4543 TCCCATTTGTCGCGGAGAGTCAAGTATGGCCAGTCACTGAGTCACTGAGTCA 4602  
QY 1460 TyrAsnValProIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1479  
Db 4603 TACATTAAGCACTGTGTGTGACAGGGCTCCACAGGGCCATCAAGAGCAAGAGTG 4662  
QY 1480 AlaValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1499  
Db 4663 GCTGTGTGACCGTGTGATGCTGTGACACAGAGTGGCTTGGCTGCGATGTGCTG 4722  
QY 1500 GlyAsnValProIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1519  
Db 4723 GGCATCTACTCTGTGTGCTGCCAGGCAACCGAGGTGACAGAGATCTTGTGATG 4782  
QY 1520 ThrGluProIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1539  
Db 4783 ACGGGGCGCTGTACTAGAGCGGGGTGCTGACCTGCGCCGAGAGCTTCCAGTCCGATG 4842  
QY 1540 ArgGlnPheValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1559  
Db 4843 CCGCAGTGTGTGTGATGAGAACTGCAAGGTGACAGCGGCACTAACAATGAGT 4902  
QY 1560 AspPheIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1579  
Db 4903 GACTTCATTTGCCAACAATGAGCACTGTGTGCTGCTCCCTGCCAAGAAGAGTGTGATC 4962  
QY 1580 SerAsnThrCysHisAsnGluValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1599  
Db 4963 ACCAACAATTTGCCAACAATGAGGCACTGTGTGAGCAAGTGTGAGAGCGCTTCAAGCTGCAAG 5022  
QY 1600 CysProIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 1619



Db 5023 TGGCCCCCTGGGCTTTGGGGGCAAGAGCTGGCCCAAGAAATGGCCAAATCCAGACACTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyr 1639  
 Db 5083 CTGGGAGAGAGCTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTCAGCTCATGTTCTGGCAGCGCCGACGAGCGGCTGCTGGTGGAGGCCAAATCCAGAG 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGly 1679  
 Db 5203 GGGCCGAGAGCAACCACTCCCTACAGCTACAGAGGGCCAGCTAGTGTGAGCCGTGAGGGCC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTrp 1699  
 Db 5263 ACAGGGCTTCAGGGCT 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPhe 1719  
 Db 5323 CACCAATGCAACACTGGCACTGGGAGCCAGCGGGGGGCTGGCCAAATGCCAATTCGTCTTC 5382  
 QY 1720 AspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisGln 1739  
 Db 5383 GATTATGGGCAAG 5442  
 QY 1740 SerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgIlePheArg 1759  
 Db 5443 AGCAACATACAGTGGGGGGGGAATCTGGGCCAGCGGGGGGTGTGGCCCTTGCTTCGG 5502  
 QY 1760 GlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspPro 1779  
 Db 5503 GGCTCTTTGGCAGGGGTGTGGGTGAGCATACGCCAGAGGGGGTTAAAGCTTGAGTCCC 5562  
 QY 1780 SerHisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSer 1799  
 Db 5563 AGCCATGGGAGAGAGATCAACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5622  
 QY 1800 AsnProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCys 1819  
 Db 5623 AACCGGTGTCCGTACAGCTATTCAGCAACGACTGGGAGACGACTATTCCTGCGAGCTGT 5682  
 QY 1820 AspProGlyTyrTyrGlyAspAsnGlyThrAsnValCysAspLeuAsnProCysGlnHis 1839  
 Db 5683 GATCCAGGTACTATGTGTACAACTGTACTAATGTGTGACCTGAACCCGCTGAGACAC 5742  
 QY 1840 GlnSerValCysThrArgIleProSerIleAspProHisGlyTyrThrCysGlnCysProPro 1859  
 Db 5743 CAGTCTGTGTGTACCCGCAAGCCCAAGTGCCTCCCAAGGCTATACCTGCGAGTGTCCCCA 5802  
 QY 1860 AsnTyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysProArgGlyTyrTrp 1879  
 Db 5803 AATTACCTTTGGGCCATACGTGTAGACAGAGATTGACAGAGCTTGTCTCCGTGGCTGGTGG 5862  
 QY 1880 GlnHisProThrCysGlyProCysAsnGlyAspValSerLeuGlyPheAspProAspCys 1899  
 Db 5863 GGAACATCCCACTGTGGCCCATGACAACTGTGATGACGCAAGGCTTTGACCCAGACATGC 5922  
 QY 1900 AsnTyrSerThrSerGlyGlnCysHisGlyAsnHisValAsnHisTyrArgProProGlySerPro 1919  
 Db 5923 AACCAAGACAGCGGCGAGTGCACATGCAAGAGAAACCACTACCGGCCCCCAGAGAGCCCC 5982  
 QY 1920 ThrCysLeuLeuCysAspCysTyrProThrArgIleSerLeuSerArgValCysAspProGln 1939  
 Db 5983 ACCTGCTCTTGTGTGTGACGTGTACCCCAAGGCTCTTGTCTCAGAGTGTGTGACCTGTAG 6042  
 QY 1940 AspGlyGlnCysProCysLeuAspProGlyValIleGlyArgGlnCysAspArgCysAspAsn 1959  
 Db 6043 GATGGCCAGTGTCAATGCAAGCCAGAGTGTCAATCGGCGCTGAGTGTGACCGCTGTGACAC 6102  
 QY 1960 ProPheAlaGlnValIleThrThrAsnGlyCysGln-----ValAsnTyrAspSer 1975  
 Db 6103 CTTTGTCTGAGGTACACACAAATGGCTGTGTGAAGGGCCCTTGTGTGTAGTTAC----- 6156

QY 1976 CysProArgAlaIleGlnAlaGlyIleTrpTrpProArgThrArg----- 1990  
 Db 6157 TGTCCCCCGCCCAATAGAGTGC-----TGGCTTCCAGCAAGAACTCTCAGCCAGTCT 6207  
 QY 1991 PheGlyLeuProAlaAlaAlaProCysProIleGlySerPhe-----Gly 2005  
 Db 6208 CAGGGGCTTCCGTGTGTGTCTCCCT---GAGGCCGGCCCTTTTGGCTTCTTCCCCCAGGG 6264  
 QY 2006 ThrAlaValAlaGlnHisCysAspGlnHisAspGlyTyrTrpLeuProProAsnLeuPheAsnCys 2025  
 Db 6265 ACTGTGTGGCCCACTGTGATGAGCAGAGGGGGTGGCTCCCCCAAACTCTTCAACTGC 6324  
 QY 2026 ThrSerIleThrPheSerGlnLeuLeuGlyPheAlaGlnGlyLeuGlnArgAsnGlySer 2045  
 Db 6325 AGCTCAATCACTTCTCAGAACTGAAAGGGCTTCCGTGAGCGGCTTACAGCGGAAATGATGTC 6384  
 QY 2046 GlyLeuAspSerGlyArgSerGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHis 2065  
 Db 6385 GGCTTAGACTCAGGGCGCTCCCAAGAGCTAGCCCTGTCTGTGGCAACGCCAGCGAGCAC 6444  
 QY 2066 ThrAlaGlyTyrPheGlySerAspValIleValAlaTyrGlnLeuAlaThrArgLeuLeu 2085  
 Db 6445 ACACTGTGCTACTTCGCGAGGAGAGAGTCAAGGTGGCTTACAGCTGGCCACGCGCTGCTG 6504  
 QY 2086 AlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThr 2105  
 Db 6505 GCCACAGAGACACCCAGCGGGGCTTTGGCTGTCTGCCACAGAGAGCTGACCTTCACT 6564  
 QY 2106 GlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsnTyrArgHisTrpGln 2125  
 Db 6565 GAGAACTGTGTGGGTGTGGGAGAGCGCTCTGAGCAACAGCAACAGCGGCACTGGAG 6624  
 QY 2126 LeuIleGlnGlnThrGlnGlnGlyTyrThrAlaTrpLeuLeuGlnHisTyrGlnAlaTyrAla 2145  
 Db 6625 CTGATTCAGAGAGAGAGAGGTGACAGCGCTGTGTCCAGCACTAGAGAGCTTACGCC 6684  
 QY 2146 SerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThrPro 2165  
 Db 6685 AGTCCCTCGGCCCAAGAACTCGGCAACCTTACCTTACGCCCTTCAACATGTCTCAGGCC 6744  
 QY 2166 AsnIleValIleSerValValArgLeuAspIleGlyAsnPheAlaGlyAlaLeuPro 2185  
 Db 6745 AACATTTGATCTCCGTAGTGTGCTTGTGACAAAGGAATTTGTGGGGGCAACTGTGCC 6804  
 QY 2186 ArgTyrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThrThrValIleLeuPro 2205  
 Db 6805 CGTACAGAGGCTGTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6864  
 QY 2206 GlnSerValPheArgGlnThrProProValValArgProAlaGlyProGlyGlnAlaGln 2225  
 Db 6865 GAGTGTGCTTCAAG 6924  
 QY 2226 GlnProGlnGlnLeuAlaArgArgGlnArgArgHisProGlnLeuSerGlnGlyAla 2245  
 Db 6925 GAGCCAGAGAGAGTGGAGCGGCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6984  
 QY 2246 ValAlaSerValIleIleTyrThrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspPro 2265  
 Db 6985 GTGGCCAGCGTCAATCACTACCGGACCTTGGCCGGGCTTACGTCAATTAATGACCTT 7044  
 QY 2266 AspTyrSerSerLeuArgValProIleArgProIleIleAsnThrProValIleSerIle 2285  
 Db 7045 GACAGAGCGAGCTTGAAGATGCCAAGAGCCCAATCAACACACCCGTGTGTAGCATTC 7104  
 QY 2286 SerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspIleProValThrValGln 2305  
 Db 7105 AGCTGCATGATGTGTGAGAGCTTCTGCCCGGGCCCTTGGACAAACCGTGTCACTGTAG 7164  
 QY 2306 PheArgLeuLeuGlnThrGlnGlnArgThrThrAspProIleCysValPheTrpAsnHisSer 2325  
 Db 7165 TTTCCGCTTGTGAG 7224



QY	2326	IIeLeuValSerGIYThnGIYTrpSerAlaArgIYSuGIuValAlaPheArgAsn	2345
Db	7225	ATCTGGTCACTGGCAACAGTGGCTGGTCCGCCAAGAGCTTGAAGTCTTCCGAAAT	7284
QY	2346	GIuSerHiSValSerCYSGInCYaasnHiEWetThSerPheAlaValLeuMetAspVal	2365
Db	7285	GAGAGCCACGTCACCTGGCTGACGTGCACCACTGACGAGCTTCGGTCTCATGAGAGCTT	7344
QY	2366	SerArgArgIu-----AanGIuIIeLeu	2374
Db	7345	TCTCGGGGGAGAGTCCGGGCCCAACAGGGGACGTGCAGAGCCGTGAAATGGGAAATCTTG	7404
QY	2375	ProLeuYstrHLeuThrTYrValAlaLeuGIYValThrLeuAlaIleuLeuLeuThr	2394
Db	7405	CCACCTGAAGACACTGACATACGTGGCTCTAGGTGTCACTTGGCTGCCCTTCTGTCAAC	7464
QY	2395	PhePhePheLeuThrHrLeuLeuArgIIeLeuArgSerAasnHiAGIYIleArgArgAsn	2414
Db	7465	TTCCTCTCTCCACCTCTTGGCGATCCGTGGCTTCCAAACAAACGGGATCCAGCTAAC	7524
QY	2415	LeuThrAlaIaIeLeuGIYLeuAlaGIuLeuValPheLeuLeuGIYIleAasnIAlaAsp	2434
Db	7525	CTGACAGTGGCCCTGGGGCTGGCTCAAGCTGGTCTTCTCTCTGGAAATCAACAGGCTGAC	7584
QY	2435	LeuProPheAlaCYStrHValIIeAlaIleLeuLeuHiSPheLeuTYrLeuCYStrH-Phe	2454
Db	7585	CTCCCTTTTGGCTGCACAGTCACTTGGCATCTCCGTGCATCTTCGTAACCTCTGCACCTTT	7644
QY	2455	SerTrpAlaLeuLeuGIuAlaLeuHiIleuTYrArgAlaLeuThnGIuValArgAspVal	2474
Db	7645	TCTCGGGCTCTGGCTGGAGGCTTGCACCTGTAAACGGGACCTCACTGAAGTCCGGGAATCTC	7704
QY	2475	AsnThrGIYProMetArgPheTYrTYrMetLeuGIYTrpGIYValProAlaPheIIeThr	2494
Db	7705	AACACCGGGCCCACTAGCGCTTCACTACATACATGCTGGGCTGGGGCGTGGCTGCCTTCAACA	7764
QY	2495	GIYLeuAlaValGIYLeuAspProGIuGIYTYrGIYAsnProAspPheCYStrHLeuSer	2514
Db	7765	GGGCTAGCGCGGGGCTGGAGCCCGAAGGGCTCAACGGAAACCTGAACTTCTGCTGGCTCTCC	7824
QY	2515	IIeTYrAspHrLeuIleTrpSerPheAlaGIYProValAlaPheAlaValSerMetSer	2534
Db	7825	ATCTAATGACAGCTCATCTGAGATTTCCTGGCCCGGCGGTGGCTTGGCGTCTGCATGAGT	7884
QY	2535	ValPheLeuTYrIIeLeuAlaAlaArgAlaSerCYaAlaAGIuArgGIuGIYPheGIu	2554
Db	7885	GCTCTTCCGTGACATCTCCGGCGGCCCTCTGTGTGGCCACAGCGGACGGGCTTTAG	7944
QY	2555	LYsIleuGIYProValSerGIYLeuGIuProSerPheAlaValIleuLeuLeuSerAla	2574
Db	7945	AAGAAAGGTCCTGTCTCGGGGCTGCAGGCTTCTTGGCCGATCTCTCTGTGTGAGGCC	8004
QY	2575	ThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspHrLeuLeuPheHISYrLeuPhe	2594
Db	8005	ACGTGGCTGTGGCACTGCTCTCTGTCAACAGCGAACCTTCTCTTCCACTCACTCTTT	8064
QY	2595	AlaThrCYaasnCYIIeGIuGIYProPheIIePheLeuSerTYrValValLeuSerIYS	2614
Db	8065	GCTACCTGCAATGTGATCCAGGGGCCCTTCACTTCTCTCTAATGTGTGTGTGAAG	8124
QY	2615	GIuValArgIYSAlaLeuIYSLeuAlaCYSerArgIYSProSerProAspProAlaLeu	2634
Db	8125	GAGGTCCGGAAAGCACTAGGTTGGCTGCAGCGCGCAAGGCCAGGCTGACCTGTCTTG	8184
QY	2635	ThrThrIYSerThrLeuThrSerSerTYrAsnCYProSerProTYrAlaAspGIYArg	2654
Db	8185	ACCAACAAAGTCAACCTCACTCGTCCCAACAATGCCACCGCCCTTAAGCAAGATGGGGG	8244
QY	2655	LeuTYrGIuProTYrGIYAspSerAlaGIYSerLeuHiIleSerThSerArgSerGIYIYS	2674
Db	8245	CTGTACCAAGCCTTACGAGAGCTCGGCCGGCTCTCTGCAACAGCAACAGTCTCGCTCGGGCAAG	8304
QY	2675	SerGIuProSerTYrIleProPheLeuLeuArgIuGIuSerAlaLeuAsnProGIYGIu	2694

Db	8305	AGTCAGCCCAAGCTACATCCCTCTTGTCTGAGGAGAGTCCGACCTGAACCTGGCCAA	83364
Qy	2695	GLYProProGlyLeuGlyAspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHis	2714
Db	8365	GGGCCCCCTGCGCTGGGGGATCCAGGAGGCTGTCTCTGAGAGGTCAAGACACAGCAT	8424
Qy	2715	AspProAspThrAspSerAspSerAspLeuSerLeuGlnuAspAspGlnSerGlySerTyr	2724
Db	8425	GATCTTCACACGGAATCCGACATGACCTCTCTTAGAAGACACAGCATGAGTGCCTCAT	8484
Qy	2735	AlaSerThrHisSerSerAspSerGluGluGluGluGluGluGluGluGluGluGluGluGlu	2754
Db	8485	GCCTCTACCATCATCATCAGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	8544
Qy	2755	PheProGlyGlyGluGlnGlyTyrPaspSerLeuLeuGlyProGlyValAspGlyLeuProLeu	2774
Db	8545	TTCCCTGAG	8604
Qy	2775	HisSerThrProIlyAspGlyGlyProGlyProGlyProGlyProGlyProGlyProGlyPro	2794
Db	8605	CACAGTACTCCCAAGATGGGGGCCAGGGGCTGGAGAGGCCCCCTGGCCAGAGAGATTT	8664
Qy	2795	GlyThrThrAlaIlyGlySerSerGlyValAspGlyAlaProGluGluGluGluGluGluGlu	2814
Db	8665	GGGACCCAGCAAGCAAGAGAGATAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT	8724
Qy	2815	GlyAspAlaLeuSerAspArgGluGlySerLeuGlyProLeuProGlySerSerAlaGlnPro	2834
Db	8725	GGAGATCCCTGTCTGAGAGAGGGGTCCCTAGGCCCCCTTCAAGGCTTTCTGCCAGCCT	8784
Qy	2835	HisIlyGlyIleLeuIlyIlyValGlyLeuProThrIleSerGluIlySerSerLeuLeu	2854
Db	8785	CACAAAGGCATCTCTTAAGAGAGATGCTCCGCCACATCAGCCAGCAAGACAGCCTCTCG	8844
Qy	2855	ArgLeuProLeuGluGluGlySerThrGlySerSerArgIlySerSerAlaSerGluGlySer	2874
Db	8845	CGGCTCCCCCTGGAGCATACACAGGGGTCTCCGGGGGCTCTCCGCTAGTGAAGGAGC	8904
Qy	2875	ArgGlyGlyProProProArgProProProArgGlnSerLeuGlnGluGlnGlnGlnGlnGly	2894
Db	8905	CGGGGCGGGCCCCCTCCCGGCCACCGCCCCGGGAGAGCCCTCCAGAGAGAGCTGAACGGG	8964
Qy	2895	ValMetProIleAlaMetSerIleValAlaGlyThrAlaAspGluAspSerSerGlySer	2914
Db	8965	GTCATGGCCATCGCCATGAGCATCAAGACAGGACAGCGATGGATGAGACTGCTCAGCTCC	9024
Qy	2915	Glu	2934
Db	9025	GAA9027	
RESULT 6			
ABA08648			
ID	ABA08648	standard; cDNA; 9401 BP.	
XX	ABA08648;		
XX	11-JAN-2002	(first entry)	
DE	Human FLAMINGO 1 homologue-encoding cDNA, SBO ID NO:424.		
XX	Human; cytokine; cell proliferation; cell differentiation; growth factor;		
KW	haematopoiesis regulation; tissue growth; immunomodulator; actvlin;		
KW	inhiblin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KW	chronic inflammatory condition; proliferative retinopathy;		
KW	atherosclerosis; coronary heart disease; arterial ischaemia;		
KW	bone disorder; osteoporosis; vascular growth disorder;		
KW	tissue regeneration; wound healing; infection; immune disorder;		
KW	cell culture; drug screening; gene therapy; antiinflammatory;		
KW	antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;		
KW	cytostatic; osteopathic; vasotrophic; cardiant; virucide; antibacterial;		

XX antifungal; vulnery; antilucer; ss.  
 XX Homo sapiens.  
 XX WO200157188-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US03800.  
 XX 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 XX (HSE-) HXSEQ INC.  
 XX Tang YF, Liu C, Drmanac RT;  
 XX MPI: 2001-457740/49.  
 XX P-FSDB; ABB11404.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 XX Claim 1, Page 496-498; 1963pb; English.  
 XX  
 XX Sequences ABB10961-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoietic regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness.  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.  
 XX  
 XX Sequence 9401 BP; 1839 A; 2961 C; 2644 G; 1957 T; 0 other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 0 Length: 9401  
 XX Score: 13378.50 Matches: 2542  
 XX Percent Similarity: 99.14% Conservative: 2  
 XX Best Local Similarity: 99.06% Mismatches: 12  
 XX Query Match: 86.06% Indels: 10  
 XX DB: 22 Gaps: 3

US-09-916-849a-3 (1-2923) x ABA08648 (1-9401)  
 QY 368 GLUAlaserARPGInGIYAARGAsPProGIYProArg-SerThrThrAlaAlaValPheLe 387  
 Db 3 GAGGCAAGTGAACAGAGGTCGGAGCCCGGGTCTCGGAGTACCAACAGCGCTGTTTCT 62  
 QY uSerValGIuAspAspAspAspAspAlaProGlnPheSerGIuYAsrGIYValAlGI 407  
 Db 63 TTCTGTGGAGGATGACATATATATGCCCCCACTTATGTGAAGAGCCATATGTGTCCA 122  
 QY nValArgGIuAspValThrProGIYAlaProValIleuArgValThrAlaSerAspArgAs 427  
 Db 123 GGTGAGGAGGATGTGACTCCAGGGGCCCGAGTCTCCAGTCAAGCTCGATTCAGAGA 182  
 QY pIyAGISeArAspAlaValAlaHisYrSerIleMeSerGIYAspAlaArgGIYGIInp 447  
 Db 183 CAAAGGAGGCAATCCGCTGGTGCATATAGCATATAGTGGCAATGCTCGGGAGACAGTT 242  
 QY eTYrIleuAspAlaGlnThrGIYAlaIleuAspValIleSerProIleuAspTYrGIuThr 467  
 Db 243 TTATCTGGATGCCAGACTGGAGCTTGAGTGTGTAGACCTTGTACTATAGAGAC 302  
 QY IYleGIuTYrThrIleuArgValArgAlaGlnAspGIYGIYArgProIleuSerAspVal 487  
 Db 303 CAAAGAGTACACCTTACGGGTGCGAGCAGAGATGTGGCCGCTCCCACTCTTAATGT 362  
 QY lSerGIYleuValThrValGlnValIleuAspIleAspAspAspAlaProIlePheVal 507  
 Db 487 CTGTGGCTTGGTACAGTACAGTCTGGAGTATCAAGACATCCCACTTGTGTGAG 422  
 QY rThrProPheGlnAlaThrValIleuGIuSerValProIleuGIYTYrIleuValIleuHis 527  
 Db 423 CACCCCTTTCAGGCTACGTGCTCGAGAGGTCCTTGGGCTTACCTGGTTCATGT 482  
 QY lGlnAlaIleAspAlaAspAlaGlyIleAspAlaArgIleuGIYTYrArgIleuAlaGIY 547  
 Db 483 CCAAGCTATGACGCTATGTGTGTGCAATGCCCGCTGGAATACCGCTGTGTGGGT 542  
 QY lGIYHisAspPheProPheThrIleAspAspGIYTYrGIYTYrIleSerValAlaAlaGI 567  
 Db 547 GAGACATGACTTCCCTTCCATCAATCAATGACAGGCTGTGATCTGTGGCTGTGA 602  
 QY uIleuAspArgGIuGIuValAspPheTYrSerPheGIYValGIuAlaArgAspHisGIY 587  
 Db 603 ACTGACCGGAGAGGAATTATTTCTACGCTTGGGGTGAAGCTCGAGCCATGGAC 662  
 QY rProAlaIleuThrAlaSerAlaSerValIleValIleuAspValIleAspAspAs 607  
 Db 587 TTCAGCACTCACTGCTCGGCTCGAGTGCAGCTGACCTGATGTCAACGACAA 722  
 QY nProThrPheThrGlnProGIYTYrThrValArgIleuAspGIuAspAlaAlaValGI 627  
 Db 723 TCCAACTTTTACCACCAAGTACAGTACAGTGGCTCAATGAGATGACGTGGGAC 782  
 QY rSerValValThrValSerAlaValAspArgAspAlaHisSerValIleThrYrGI 647  
 Db 627 CAGGTGTGACGGGTGCAGCTGTGACCGTGAATCATATGTCACTTACACGAT 842  
 QY eThrSerGIYAsnThrArgAspArgPheSerIleThrSerGISeSerIYGIYGIYleuVal 667  
 Db 647 CACCAAGTGAATATCTGAAACCGCTTCTCAATCAACGCAAGGTGGTGGGCTGT 902  
 QY lSerIleuAlaIleuProIleuAspTYrIYleuGIuArgGIYTYrValIleuAlaValThrAl 687  
 Db 903 ATCCCTTGCCCTGCACTGACCTACAACTTGACCGGCACTATGTGTGGCTGTACCGC 962  
 QY aSerAspGIYThrArgGlnAspThrAlaGlnIleValAlaAsnValThrAspAlaSer 707  
 Db 687 CTCCGATGAGCACTCGGAGGACAGGACAGATGTGTGTAATGTCAACGCAAC 1022  
 QY rHisArgProValPheGlnSerSerHisTYrThrValAsnValAsnGIuAspArgProAl 727

Db 1023 CCATGCTCTGCTTTACAGCTCCCACTATACAGTAATGTTATATGAGACCGGCGGC 1082  
Qy 727 aglyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgI 747  
Db 1083 AGGCACACAGGTGTGTCTGATCAGCGCCAGATGAGACACAGGTGAGAAATCCCGCAT 1142  
Qy 747 ethrTyrrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaVala 767  
Db 1143 CACCTTCCTTCATGAGACAGCATCCCGCAGTTCGCCATCATGATGACAGACACGGGGCTGT 1202  
Qy 767 IThrThrGlnAlaGluLeuAspTyrrGluAspGlnValSerTyrrThrLeuAlaIleThrAl 787  
Db 1203 CACCAACCAAGCTGACGTGACCTACAGACACAGAGTCTTACACCTCGGCATTTACTGC 1262  
Qy 787 aArgAspAsnGlyIleProGlnIleSerAspThrTyrrLeuGlnIleLeuValAsnAs 807  
Db 1263 TCGGAGCAATAGGCAATCCCAAGAGTCCGACACACTACCTGAGAAATCTGTGTAGACA 1322  
Qy 807 pValIAsnAspAsnAlaProGlnPheLeuArgAspSerTyrrGlnGlySerValTyrrGluAs 827  
Db 1323 CGTGAATGACAAATGCCCCCTCAGTTCCTGCGAGACTCTACCAAGGGCAGTGTCTATAGAGA 1382  
Qy 827 pValProProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAs 847  
Db 1383 TGTGCCACCTTCCTACAGGCTCTGAGATCTCAGCCACTGATCGATTCGAGACTTAA 1442  
Qy 847 nglyArgValIlePheTyrrThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValG 867  
Db 1443 TGGCAGAGGCTTCTTCAACCTTCCAGAGGCGACGATGAGACGGTAACTTTATGTGTA 1502  
Qy 867 userThrSerGlyIleValIleArgThrIleuArgArgLeuAspArgGluAsnValAlaGlnTy 887  
Db 1503 GTCCAGGTACAGGCTGTGCGAAGCGCTACGAGGCTGATCGAGAAAGGTGGCCACAGTA 1562  
Qy 887 rValIleuArgAlaTyrrAlaValAspIleGlyMetProProAlaArgThrProMetGluVala 907  
Db 1563 TGTCTTGGCGGCAATATGACGTGACAGAGGAGATGCCCAAGCCCGACACTATGAGAAAT 1622  
Qy 907 IThrValIThrValIleuAspValIAsnAspAsnProProValIlePheGlnIleAspGluPheAs 927  
Db 1623 GACAGTCACGTGTGTGATGTGAAATGACATCCCGTCTTGTAGACAGATGAGTTGTA 1682  
Qy 927 pValIleuValIleGluIleuAspSerProIleGlyLeuAlaValAlaArgValIThrAlaThrAs 947  
Db 1683 TGTGTTTGGAGAGAAACAGCCCAATGGCTAGCGGTGGCCGGGTACACGCCACTGA 1742  
Qy 947 pProAspGluGlyThrAsnAlaGlnIleMetTyrrGlnIleValIleGluGlyAsnIleProG 967  
Db 1743 CCCCAGTGAAGGACCAATGCCCAAGTATGTAACAGATTGTGAGGGGCAACATCCCTGA 1802  
Qy 967 uValIleuGlnLeuAspIlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrrG 987  
Db 1803 GGTCTTCAGCTGAGCATCTTCTCCGGGAGGTGACAGCCCTGTGTAGACTTATGACTACGA 1862  
Qy 987 uAspArgProGluTyrrValIleuValIleGlnAlaThrSerAlaProLeuValIleSerArg 1007  
Db 1863 GACCGGCTGTAGTACGTCTGTGTATCCAGGCAAGTCAAGCTCTCTGTGTAGACCGGCGC 1922  
Qy 1007 aThrValIleValIleArgIleuLeuAspArgAspAsnAspProProValIleuGlyAsnPheG 1027  
Db 1923 TACAGTCCACGTCCGCTCTTGAACCGCAATGACACCAACAGTGTCTGGGCAACTTTGA 1982  
Qy 1027 uIleLeuPheAsnAsnTyrrValIThrAsnArgSerSerSerPheProGlyGlyAlaIleG 1047  
Db 1983 GATCCCTTTCACAACTATGTCTACCAATGCTCAACACACTTCCCTGGGGGTGCCATTTGG 2042  
Qy 1047 yArgValIProAlaIleAspProAspIleSerAspSerLeuThrTyrrSerPheGluAspG 1067  
Db 2043 CCGAGTACCTGCGCAAGACCCGTGATATCTCAGATATCTGACTTACAGCTTTGAGGGGGG 2102  
Qy 1067 yAsnGluLeuSerLeuValIleuLeuAsnAlaSerThrGlyGluLeuValLeuSerArgAl 1087  
Db 2103 AATGAACTCAGCTGTGTCTGTCAATGCTTCAAGGGTGAAGCTGAAGCTGAAGCGGCGC 2162

Qy 1087 aLeuAspAsnAsnArgProLeuGlnAlaIleMetSerValIleuValIleAspGlyValAlaI 1107  
Db 2163 ACTGACAAACAACCGGCTCTGAGGGCCATATGAGCGGTGTGTGACAGCGGCTACA 2222  
Qy 1107 sSerValIThrAlaGlnCyAlaLeuArgValIThrIleIleThrAspGluMetLeuThrHi 1127  
Db 2223 CAGCGTGAACCGCCGAGTCCGGCTGTGTGACCAATCATCAACCGATGAGATGTCTTACCA 2282  
Qy 1127 sSerIleThrLeuArgLeuGluAspMetSerProGluArgPheLeuSerProLeuLeuG 1147  
Db 2283 CAGCATACGCTGCGCTGTGAGACATATCAACCGAGCGCTTCTCTGATCAACTCTAGG 2342  
Qy 1147 yLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrProProAspAlaValValPh 1167  
Db 2343 CCTTTATTCACAGGGGGGTGGCCGCGCACGCTGGCCACCGCACCGAGTGTGTCTT 2402  
Qy 1167 eAsnValGlnArgAspThrAspAlaProGlyGlyHisIleIleLeuAsnValIleLeuSerVal 1187  
Db 2403 CAACGTACAGCGGGAACACCAAGCGCCCGGGGGCCAACTCTCAACGTGAGCTGTGCGT 2462  
Qy 1187 lGlyIleProProGlyProGlyGlyGlyProProPheLeuProSerGluAspLeuGlnI 1207  
Db 2463 GGGCTACGCGCAAGGCGCCGGGGCGGGCGCGCTTCTGCTGCTGTAGAGACTGTAGGA 2522  
Qy 1207 uArgLeuTyrrLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValIleuProPh 1227  
Db 2523 GCGCTTATTCCTCAACCGCACGCTGTGACGGCCATCTCGGACACAGCGGTGTGCTCTT 2582  
Qy 1227 eAspAspAsnIleCyAlaLeuArgIleProCyArgIleuAsnTyrrMetArgCyValIleSerVal 1247  
Db 2583 GACACCAACAATCTGCTGTGGAGGCTCTGTGAGAACTCAATGCGCTGTGTGCGTCT 2642  
Qy 1247 uArgPheAspSerSerAlaProPheIleAlaSerSerSerValIleuPheArgProIleHi 1267  
Db 2643 GGGCTTCACTCTCTCGGCGCTTATATGCTCTCTCTCGTGTGCTCTTCCGGCCATCA 2702  
Qy 1267 aProValIleGlyLeuArgCyValArgCyAspProGlyPheThrGlyAspTyrrCyArgIleTh 1287  
Db 2703 CCCCCTCGAGAGGCTGCGCTCCGCTGCGCCCGCGCTTCAACGGGTGACTACTGTGAGAAC 2762  
Qy 1287 rGluValIleuLeuCyAspTyrrSerArgProCyArgIleProHiIleGlyArgCyAspSerArg 1307  
Db 2763 CAGGTGAGACCTGTGCTACTGTGCGGCTGTGTGAGCCCAACGGAGCTGTCCGACCGCGGA 2822  
Qy 1307 uGlyGlyTyrrThrCyAspLeuCyAspAspGlyTyrrThrGlyGluHisCyArgIleuValIleSer 1327  
Db 2823 GGGGGGCTACACTGTGCTCTGTGTGTGATGCTACACGGGTGAGCACTGTGAGTGTAGTGC 2882  
Qy 1327 aArgSerGlyArgCyThrProGlyValCyValAsnGlyGlyTyrrCyValAsnLeuLe 1347  
Db 2883 TCGCTCAGGCGGTGCAACCCGGGTGTCTGTGCAAAATGGGGGCACTGTGTCAACTTCT 2942  
Qy 1347 uValIleGlyPheIleCyAspAspCyProSerGlyAspPheGlyIleuTyrrCyArgIleVal 1367  
Db 2943 GGTGGGCGGTTCAGATGCGATGTCCCATCTGAGACTTCAAGAAAGCTTACTGCGAGGT 3002  
Qy 1367 IThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPh 1387  
Db 3003 GACCAACGGGAGCTTCCCGCCCACTCTTCATCACTTGTGGGCGCTGCGCCAGCGTTT 3062  
Qy 1387 eHisPheThrIleuAlaLeuSerPheAlaThrTyrrGlyIleArgAspGlyLeuLeuLeuTyrrAs 1407  
Db 3063 CCACTTCAACCTGCGCTCTGTGTGTTCACAAAGAGGCGCAACGGGTGTGTGTATCAA 3122  
Qy 1407 nglyArgPheAsnGluTyrrHisAspPheValAlaLeuGlnValIleGlnGlnGlnValIleG 1427  
Db 3123 TGGGCTTTCATAGAGAGCATGACTTTGTGGCCCTGAGAGTATCCAGAGCAGCTGTCA 3182  
Qy 1427 nLeuThrPheSerAlaGlyIleuSerThrThrValIleSerProPheValProGlyGlyVala 1447  
Db 3183 GCTCACTTCTCTGAGAGGAGGTCAACCAACAGGTGTGCCCAATTCGTGCGCGGAGGAGT 3242

QY	1447	SerzAspGlyGlnTrpHisSerThrValGlnLeuLeuSerThrTyraHisLeuProLeuLeuGlyGly	1467
Db	3243	CAGTATGAGCCAGGTGGCAATACGGTGGACACTGAAATCAATACATATAGACCACTGTGGGTGC	33020
QY	1467	nThrGlyLeuProGlnGlyProSerGlyGlnTyraValAlaValAlaThrValAspGlyCys	1487
Db	3303	GACAGGGCTCCCAACAGGGCCCACTCAAGACGAGAGGTGGCTGTGGTACCGGTGATGGCTG	3362
QY	1487	aAspThrGlyValAlaAlaLeuArgPheGlySerValLeuGlyValnTrpSerCysAlaAla-G	1507
Db	3363	TGACACAGAGGTGGCTTGCGCTTGCGATCTGTCTCTGGGCAACTACTCTGTGGTGGCC	3422
QY	1507	InglyThrGlnGlyGlySerTyraValSerLeuAspLeuThrGlyProLeuLeuLeuGlyG	1527
Db	3423	AGGGGACCAAGGTGTGGCAAGAAAGTCTGTGATCTCAACGGGGCCCTGCTACTAGGCG	3482
QY	1527	lyValProAspLeuProGlyIleProSerPheProValArgMetArgGlnPheValGlyCysMet	1547
Db	3483	GGGTCCCTGACCTTGCCCAAGACTTCCAGTCCGATCGGAGGGAGTGTCTGGGCTGCATGC	3542
QY	1547	rgAsnLeuGlnValAspSerArgHisIleAspMetAlaAspPheIleAlaAsnAnglyLT	1567
Db	3543	GGAACCTCGAGGTGGACAGCGCGGCACTAGACATGGCTGACTTCACTTCCACAAATGGCA	3602
QY	1567	hTrValProGlyCysProAlaTyraValSerValnValCysAspSerAsnThrCysHisAsnGlyG	1587
Db	3603	CCGTGCTCGGTGGCTCCCTCCCAAGAAAGACGTGTGACAGCAAACTTGCCACATGGGG	3662
QY	1587	lyThrCysValAsnGlyTrpTrpSerAlaPheSerCysGlyIleCysProLeuGlyPheGlyGlyL	1607
Db	3663	GCACTTGCGTGAAACAGTGGAGCGGTTCAGCTGTGAGATGTGCCCCCTTGAGCTTTGGGGGCA	3722
QY	1607	ysSerCysAlaGlnGlnLeuMetAlaAsnProGlnHisPheLeuGlySerSerLeuValAlat	1627
Db	3723	AGAGCTGGGCCCCAGGAATATGGCCATTCACAGCACTTCTGGGAGAGAGCTGTGGGCTT	3782
QY	1627	TrpHisGlyLeuSerLeuProIleSerGlnProTrpTyraLeuSerLeuMetPheArgThr	1647
Db	3783	GGCAATGGCTCTCGCTGGCCCACTCTCCCAACCTGTGATCTCAAGCTTCACTGTTCGGACGC	3842
QY	1647	rgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuG	1667
Db	3843	GCAGAGCCGACGGGTGCTCGCTGCAGGCGCATCCACAGGAGGGGCGAGCACTACCCCTAC	3902
QY	1667	InLeuArgGlyGlyHisValMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLT	1687
Db	3903	AGCTACAGAGAGGCCACCTGATCTGACGCTGAGAGGACACAGGCTTACAGGCTCTCTTC	3962
QY	1687	euarLeuGlnProGlyValArgAlaAsnAspGlyAspTrpHisHisAlaGlnLeuAlaLeuG	1707
Db	3963	TCCGCTCTGAGGCAAGGCGGGCCATATGACGGGTGACATGCAACATGCAACAGCTGGGACCTGG	4022
QY	1707	lyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAspTyraGlyGlnGlnArgAlaG	1727
Db	4023	GAGGCATCGGGGGGCTGGCCCATGCACTTGTCTTGATTAATGGGACGACGAGAGCG	4082
QY	1727	InglyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyGlyLT	1747
Db	4083	AGGGACACTGGGCCCCCGGGCTGCATGGTCTGCACCTAGCAACATACATAGTGGCGGAA	4142
QY	1747	IeProGlyProAlaGlyGlyValAlaAsnArgGlyPheArgGlyCysLeuGlnGlyValArgV	1767
Db	4143	TACCTGGGCGACGCGCGGTGTGGCCCGTGGCTTTCCGGGGCTGTGTGACGGGTGTGGGG	4202
QY	1767	AlSerAspThrProGlyGlyValAlaAsnSerLeuAspProSerHisGlyGlySerIleAsnV	1787
Db	4203	TGAGGGAATACGCACAGGGGGGTAAACACTGTGAATCCAGGCCATGGGGAGAGCATCAAG	4262
QY	1787	AlGlyGlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerLT	1807
Db	4263	TGAGCAAGGCTGTAGCTGTGACCTTGTGATCTCAAAACCGGTCTCTTACAGCT	4322
QY	1807	yrCysSerAsnAspTrpAspSerTyraSerCysSerCysAspProGlyTyraTyraGlyAspA	1827

Db	4353	ATTGAGACAAAGACTGGGACAGCTATTCTCGACACTGGATCCAGGTTACTATGCTGACAA	4382
Qy	1827	snCy5ThrAspValCyAspPLeuAspProCysgluHISgluSerValCysThrArgLysP	1847
Db	4383	ACTGTACTAAATGTGTGTGACTTGAACCCGTGTGTAGACACAGCTCTGTGTGTACCCGCAAGC	4442
Qy	1847	roSerAlaProHISglYThrThCysgluCysProProAsnTyrLeuGIYProTyrCysG	1867
Db	4443	CCAGTGCCCCCAATGGCTATACCTGGAGGTGTCCCCCAATTATCTTGGGCAATACGTGTG	4502
Qy	1867	luthArgILeAspGlnProCysPProArgglYTPTrpGlyHISProThrCysglYProC	1887
Db	4503	AGACCAAGATTTGACCAAGCTTGTGTCCTGGCTGTGTGGGACATCCCAAGATGGCCCAT	4562
Qy	1887	ysAsnCyAspValSerTyrsgLYpheAspProAspCyAsnLysThrSecrIYglLucySH	1907
Db	4563	GCACTGTGATGTCCAGCAAGGCTTTGACCCAGACTGCACAAACAGCAAGGCGGACAGTGC	4622
Qy	1907	ISCyLysgluAsnHISYrZArgProProglYserProThnCysleuLeuCysAspCysT	1927
Db	4623	ACTGGAAGAGAACCACTACCGGCCCCCAGGCAAGCCCACTGGCTCTTGTGTGACGTGCT	4682
Qy	1927	YrProThngLYSerLeuSerArgValCyAspProgluAspGlyGlnCysProCysLysP	1947
Db	4683	ACCACCAAGGCTCTTGTGCCAGAGTCTGTGACCTCGAGAGATGGCAGATGTCCATGCAAGC	4742
Qy	1947	roglYValIIleGlyArgGlnCysAspArgCysAspAspProPheAlaGluValThrThra	1967
Db	4743	CAGGTGTATGGGCGGTACGTACGTACCGCTGTGACAACTTTTGTGTAGGTTGACCAACA	4802
Qy	1967	snGlyCySgluValAsnTyrAspSerCysPProArgAlaIleGluAlaGlyIleTyrTrpP	1987
Db	4803	ATGGTGTGAAGTGAATTATATGACAGCTGGCCACAGCGATTTAGAGCTGGATCTGTGTGC	4862
Qy	1987	roArgThrArgPheGlyLeuProAlaAlaAlaProCysProLysGlySerPheGlyThra	2007
Db	4863	CCCGTACCCGCTTGGGCGCTGCTCTGCTGCTCTGCTCCAAAGGCTCTTGGGACTG	4922
Qy	2007	laValArgHISCyAspAspGluHISArgGlyTyrLeuProProAsnLeuPheAsnCyAspThrs	2027
Db	4923	CTGTGCGCACTGTGATGAGCAAGGGGGTGGCTCCCCCAACCTCTTCAACTGACGT	4982
Qy	2027	erIleThrPheSerGlnleuLysglYpheAlaGluArgLeuGlnArgAsnLysSerGlyL	2047
Db	4983	CCATCACTTCTCAGAACTGAAGGGCTTCCGTGAGCGGCTTACAGCGGAATAGTCAAGGCC	5042
Qy	2047	euAspSerGlyAspSerGlnGlnleuAlaleuLeuLeuAspAsnAlaThrGlnHISThra	2067
Db	5043	TAGACTCAGGCGGCTCCACGACGACTAACCTGCTCTGCGCAAGCGACGACGACAG	5102
Qy	2067	laGlyYrPheGlySerAspValIlyValAlaIlyrGlnleuAlaThrArgLeuLeuAlaH	2087
Db	5103	CTGGTACTTCTGGGAGGAGCTCAAGGTGGCTTACAGAGCTGGCCAGCGGCTGTGGCCC	5162
Qy	2087	laGluSerThrGlnArgglYpheGlyLeuSerAlaThrGlnAspValHISpheThrGlnAla	2107
Db	5163	ACGAGAGCAACCAAGCGGGCTTTGGGCTGTCTGCGACACAGAGCTGCACTTCACTAGA	5222
Qy	2107	snLeuLeuArgValGlySerAlaleuLeuAspThrAlaAsnLysArgHISTrpGlnleuL	2127
Db	5223	ATCTGTGCGGGTGGGAGGCGCTCTCTGGACACAGCAACAAGCGGCACTGGGAGCTGA	5282
Qy	2127	leGlnGlnThrGlnGlyGlyThrAlaThrLeuLeuGlnHISYrGlyGlnAlaIlyrAlaSerA	2147
Db	5283	TTCCAGCAAGACAGAGGTGGACCGCGCTGGCTCTCCAGCACTAGAGGCGCTTACCGCAGTG	5342
Qy	2147	laLeuAlaGlnAspMetArgHISThrTyrLeuSerProPheThrIIleValThrProAsnI	2167
Db	5343	CCCTGGCCCAAGAAATGGCGCACACTTACCTAAGCCCTTACCAATGTGTACAGCGCCACA	5402
Qy	2167	leValIIISerValValArgLeuAspLysglYAsnPheAlaGlyValAlaLysLeuProArgT	2187



QY 2902 elyvalaglythrvalaspgluaspsergylsrglupheleuphepnehele 2922  
 DB 7623 CAGGAGGACCGGTGATGAGGATCGTCAGGCTCCGATTTCTTCTTAACCTTCT 7682  
 QY 2922 whis 2923  
 DB 7683 GCAT 7686

RESULT 7  
 AAC76401  
 ID AAC76401 standard; cDNA; 9121 BP.  
 AC AAC76401;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF1956 polynucleotide sequence SRQ ID NO:3911.  
 XX

Human; open reading frame; ORF; detection; cytostatic; hepatotropic;  
 vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
 antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 cholesterol ester storage; systemic lupus erythematosus; infection;  
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 thrombosis; contraceptive; ss.

XX  
 XX Homo sapiens.  
 XX  
 XX MO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 XX  
 PR 02-APR-1999; 99US-0127636.  
 XX  
 PR 05-APR-1999; 99US-0127728.  
 XX  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkens RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 XX  
 P-PsDB; AAB42192.  
 XX

Novel nucleic acids and peptides derived from open reading frame X,  
 useful for treating e.g. cancers, proliferative disorders,  
 neurodegenerative disorders and cardiovascular disease -

Claim 5; Page 3061-3067; 5507pp; English.

AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 which represent the human ORF open reading frames 1 to 3161. The ORF  
 sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 osteopathic; immunostimulant; antiparkinsonian; antiproliferative;  
 immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 antidiabetic; hypotensive; dermatological; immunosuppressive;  
 antiinflammatory; antibacterial; antiviral; antineumatic; antithyroid;  
 antianemic. The sequences can be used for determining  
 the presence of or predisposition to, or preventing or treating  
 pathological conditions associated with an ORF-associated disorder. The  
 nucleic acids can be used to express ORF proteins in gene therapy  
 vectors. The proteins and nucleic acids may be used to treat cancers,  
 proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX

SQ Sequence 9121 BP; 1797 A; 2892 C; 2529 G; 1901 T; 2 other;

Alignment Scores:  
 Pred. No.: 0 Length: 9121  
 Score: 12839.50 Matches: 2413  
 Percent Similarity: 98.74% Conservative: 23  
 Best Local Similarity: 97.81% Mismatches: 28  
 Query Match: 82.60% Indels: 3  
 DB: Gaps: 2

US-09-916-849a-3 (1-2923) x AAC76401 (1-9121)

QY 457 AspValIserProLeuAspTyrGluThrThrIysGlyTyrThrLeuArgValArgAla 476  
 DB 4 GATGTGGTACCCCTTGTACTATGAGCAGACCAAGGATACCTTACGGGTGGACCA 63  
 QY 477 GlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeu 496  
 DB 64 CAGATGTGGTGGCCGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123  
 QY 497 AspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlu 516  
 DB 124 GATATCAACGACCAATGCCCCCATCTTCGTCCAGACCCCTTCCAGGCTACGTCCTGAG 183  
 QY 517 SerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaGlyAsp 536  
 DB 184 AGCTGCCCTTAAAGCTACCTGTTCTTCATCTCAAGCTATGACAGCTGATCTGTGAC 243  
 QY 537 AsnAlaArgLeuGlyTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsn 556  
 DB 244 AATGCCCGCTGAAATACCGCTTGTGGGTGGGACATGATCTCCCTTACCATCAAC 303  
 QY 557 AsnGlyThrGlyTyrPylIserValAlaAlaGlyLeuAspArgGlyGluValAspPheTyr 576  
 DB 304 AATGACACAGGCTGATCTCTGTGGCTGTGACCTGACCGGAGGAGATTTTCTAC 363  
 QY 577 SerPheGlyValGlyAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerVal 596  
 DB 364 AGCTTGGGGTGAAGCTCGAGACCATGCACTCAGACCTCATCTGCTGGCCAGTCTC 423  
 QY 597 SerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnProGlyTyrThr 616  
 DB 424 AGCGTACCTGCTCGATGTCACGACCAACATCCATCTTACCCCAACGAGATGACCA 483  
 QY 617 ValArgLeuAsnGlyAspAlaAlaValGlyThrSerValValThrValSerAlaValAsp 636  
 DB 484 GTGGGGCTCAATGAGGATGCGCTGTGGCAACGCGTGGAGCGGTGACGTGTGAC 543  
 QY 637 ArgAspAlaHisSerValIleThrThrGlnIleThrSerGlyAsnThrArgAsnArgPhe 656  
 DB 544 CGTATGCTCATATGTCATCATCACTTACCATGCACTGCAATCTCGAATCCGCTTC 603  
 QY 657 SerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIys 676  
 DB 604 TCCATTCACCAAGCCAAATGGGTGGGTGGCTGTGATCCCTTGGCCCTGCACTGACCAAA 663  
 QY 677 LeuGluArgGlyGlyTyrValLeuAlaValThrAlaSerAspGlyTyrArgGlnAspThrAla 696  
 DB 664 CTGAGCGGCGATGATGATGCTGTATCCGCTCCAGTGGCACTGGCGAGACCGGCA 723  
 QY 697 GlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHis 716  
 DB 724 CAGATTGGTGATGATGACCGACGCAACCAATGCTGTCTTCTTCAAGAGCTCCAC 783  
 QY 717 TyrThrValAsnValAsnGlyAspArgProAlaGlyThrThrValValLeuIleSerAla 736



Db	784	TATACAGTGAATGTTAATAGAGGACCGGCGGACGACCAAGGTGGTCTGATACGGCC	843
Qy	737	ThraSpGluAapThrGlyGluAaAlaArgIleThrTyPheMetGluAapSerIlePro	756
Db	844	ACGATATGAGGACACAGGTGAGAAATGCCCGATCACTTACATTGAGGAGACAGATCCCC	903
Qy	757	GlnPheAArgIleAaPAlaAaPThGlyValAlaValThrThrGlnAlaGluAaPThGly	776
Db	904	CAGTTCCGATCGATGACAGACAGGGGGCTGTCAACCAAGGCGGTGAGCTGAACTTACAGA	963
Qy	777	AaPAlaValSerTyPThrLeuAlaIleThrAlaArgAaPangiYIleProGlnAysSer	796
Db	964	GACCAAGGTCTTACACCTCGGCATTATCTCTCGGAGCAATGGCAATTCCCAAGATCC	1023
Qy	797	AaPThrThrTyPThrLeuGlnIleLeuValAaAaPValAaAaPAlaAlaProGlnPheLeu	816
Db	1024	GACACCACTTACCTGGAGATCTGTGTGAACAGATGAATGCAATGGCCCTTCAGTTCTGTG	1083
Qy	817	ArgAaPSerTyPThrGlnAysSerValTyPThrGluAaPAlaProPheThrSerValLeuGln	836
Db	1084	CGAAGACTCTTACAGGGGCAGTGTCTATGAGGATGTGGCACTTCACTAGGGTCTTGCAG	1143
Qy	837	IleSerAlaThrAaPArgAaPSerGlyLeuAaGlnGlyArgValPheTyPThrPheGlnGly	856
Db	1144	ATCTCAGCGCATCTGATCTGTGATCTTGACATTATGTGCAAGGGGTCTTTCACCTTCAAGGA	1203
Qy	857	GlyAaAaPArgIyAaPArgIyAaPPhaIleValGluSerThrSerGlyIleValArgThrLeu	876
Db	1204	GCGCAGATGGAGACCGGTGACTTATATGTTATGTAGTCCAGTCAAGGCAATGTGCAACGCTA	1263
Qy	877	ArgArgLeuAaPArgIyAaAaPValAlaGlnTyPValLeuAaArgAlaTyPAlaValAaPlys	896
Db	1264	CGGAGGCTGATCTGAGAGAAAGTGGCCAGATGTCTTGGCGGCAATGACAGTGAACAAG	1323
Qy	897	GlyMetProProAlaArgThrPheMetGlnValThrValThrValIleAaAaPValAaAaP	916
Db	1324	GGAATGCCCCCAGCGCCGACACTTATGGAAGTGAACAGTCACTGTGTGGAATGTAAATGAC	1383
Qy	917	AaAaPProProValPheGluGlnAaPAluPheAaPValPheValGluGluAaAaPProIle	936
Db	1384	AATCCCCCTGTCTTTGAGCAGGAATGAGTTGATGTGTGTGGAAGAAACAGCCCCATT	1443
Qy	937	GlyIleuAlaValAlaArgValThrAlaThrAaPProAaPArgIyGlyThrAaAlaGlnIle	956
Db	1444	GGGCTAGCCGCGGCGCCGGGTCAACGCCCTGACCCCAATGAAGCAACCAATGCCCAAGTT	1503
Qy	957	MetTyPThrGlnIleValGlnGlyAaMetIleProGlnValPheGlnLeuAaPAlaPheSerGly	976
Db	1504	ATGTAACCAATGTGTGAGGGGCAACATCCCTTAAGGTCTTCAAGCTGGAATCTTCTCGGG	1563
Qy	977	GluIleuThrAlaLeuValAaPLeuAaPThyPThrAaAaPArgProGluTyPValLeuValIle	996
Db	1564	GAGCTGACAGCCCTGTGTGACTTGTAGACTTACAGAGGACCGGCTGTGATAGTCTGTGGTATC	1623
Qy	997	GlnAlaThrSerAlaProLeuValSerProAlaThrValIleValArgLeuLeuAaPArg	1016
Db	1624	CAGGCCAGCTGAGCTTCTGTGTGAGCGGGCTTACAGTCCAGCTCCCTTGAACGCC	1683
Qy	1017	AaAaPAlaProProValLeuGlyAaAaPheGlnIleLeuPheAaAaPThyPValThrAaAa	1036
Db	1684	AATGACAAACCAACGAGTGTGGGCAACTTTGAGATCTTTTCAACAACTATGTCAACAAAT	1743
Qy	1037	ArgSerSerSerPheProGlyGlyValIleGlyArgValProAlaIleAaPProAaPArgIle	1056
Db	1744	CGCTCAACGAGGCTTCTCTGGGGGTCCAAATGGCCGAGTACCTGTGCCAATGACCCCTGAATAC	1803
Qy	1057	SerAaPSerLeuThrTyPThrPheGlnArgIyAaGlnGluLeuSerLeuValLeuLeuAaAa	1076
Db	1804	TCAGATATGCTGACTTACACGCTTTGAGCGGGGAATGAATCACTCAGCTGTGTCTTGCAT	1863
Qy	1077	AlaSerThrGlyGluLeuValLeuSerArgAlaLeuAaPAlaAaPArgProLeuGlnAla	1096

[illegible]

1457 LeuLYSTYTYRAsnLYsPProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlu 1476  
2995 CTGAATATCTACAAATAGCCACTGTGGGCTCAGACAGGGGCTCCACAGGGCCATCAGAG 3054  
1477 GlnLYsValAlaValValThrValaerGlyCyAspThrGlyValAlaLeuArgheGly 1496  
3055 CAGAAAGTGGCTGTGGTGAACCGTGAATGGCTGTGAACAAGAGTGGCTTGGCTTGGGA 3114  
1497 SerValLeuGlyAsnLYsSerCyAspAlaGlnGlyThrGlnGlyGlySerLYsLYsSer 1516  
3115 TCTGCTCTGGGCACTATCTCTGTCTGTCCAGGGCACTCAGGGTGGCAGCAAGAGTCT 3174  
1517 LeuAspLeuThrGlyProLeuLeuGlyGlyValProAspLeuProGlyLeuSerPro 1536  
3175 CTGGATCTGACGGGGGCTCTGCTCACTAGGCGGGGTGCTGACTGGCCCGAGAGCTTCCA 3234  
1537 ValArgMetArgGlnPheValGlyCyMetArgAsnLeuGlnValAspSerArgHisLe 1556  
3235 GTCCCAATGGCGGCACTTGTGGGCTGTGACATGGCAACCTGCAGTGGACAGCGGCACTA 3294  
1557 AspMetAlaAspPheLeuAlaAsnAsnGlyThrValProGlyCyAspProAlaLYsLYsAsn 1576  
3295 GACATGGCTGACTTATCTCCAAATGGCACTGTGCTGGCTGGCTCTGCCAAGAGAGAC 3354  
1577 ValCyAspSerAsnThrCyAspHisAsnGlyGlyThrCyAspValAsnGlnTrpAspAlaPhe 1596  
3355 GTGTGTGACAGCAACACTTGCACATGGGGGCACTTGGGTGAACCAATGGGAGCGGTTTC 3414  
1597 SerCyAspGlyCyAspProLeuGlyPheGlyGlyLYsSerCyAspAlaGlnLeuMetAlaAsnPro 1616  
3415 AGCTCGAGTGTGCTCTGGGCTTGTGGGCAAGACTGCGCCAGAGAAATGGCCCAATCCA 3474  
1617 GlnHisPheLeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProHisSerGln 1636  
3475 CAGCACTTCTCTGGGAGAGACCTGTGGCTGTGGCACTGGCTCTCTGCTCCATCTCCAA 3534  
1637 ProTrpLYsLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAla 1656  
3535 CCTGTGTACTCAGGCTCATGTTCCGACAGCGCCAGCGGCGGCTGTGCTGTGACAGGCC 3594  
1657 IleThrArgGlyArgSerThrLeuThrLeuGlnLeuArgGlnGlyHisValMetLeuSer 1676  
3595 ATCACAGAGGGGCGGAGCACCACTACCTCAAGCTCAAGAGGGCCAGCTGATGTGAGC 3654  
1677 ValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAsp 1696  
3655 GTGGAGGGCAAGAGGCTTCAAGGCTCTCTCTCTGCTGTGGAGCCAGGCCGGGCCAAATGAC 3714  
1697 GlyAspTrpHisHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaLe 1716  
3715 GGTCGCTGGCACCAATGACAGCTGGGCACTGGGAGCCAGCGGGGGCTGTGCCATGCCATT 3774  
1717 LeuSerPheAspTrpGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGly 1736  
3775 CTGTCTCTGTGATTAAGGGCAGAGAGAGAGAGGCAACTGGGGCCCCGCTGTGATGT 3834  
1737 LeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArg 1756  
3835 CTGCACCTGAGCAACAATACAGTGGCGGAATACCTGGGCCAGCGGGGGGTGTGGCCCT 3894  
1757 GlyPheArgGlyCyAspLeuGlnGlyValArgValSerAspTrpProGlnGlyValAsnSer 1776  
3895 GGCCTTCCGGGCTGTTTTCAGGGTGTGGCGGTGAACGATACGCCAGAGGGGGTTAAACAGC 3954  
1777 LeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCySerLeuProAspPro 1796  
3955 CTGAATCCCAAGCATGGAGAGAGATCAACGTGGAGAAAGGTGTGAGCTGTGACCTT 4014  
1797 CyAspSerAsnProCyAspProAlaAsnSerLYsSerAspAspTrpAspSerLYsSer 1816  
4015 TGTGACTCAAAACCGGTGTCTGTCTAAAGCTATTTCAGCAACGACTGGGAGACGCTATTCC 4074

1817 CysSerCyAspProGlyLYsTrpGlyAspAsnCyThrAsnValCyAspLeuAsnPro 1836  
4075 TGCAGCTGTGATCCAGGTTACTATGTGTGACAACTGTACTAATGTGTGACTGAAACCG 4134  
1837 CysGlnHisGlnSerValCyThrArgLYsProSerAlaProHisGlyLYsThrCyAspGlu 1856  
4135 TGTGAGCACCAAGTGTGTGTACCCCGCAAGCCAGTGTGCCCCCAATGGCTAATACCGCAG 4194  
1857 CysProProAsnTrpLeuGlyProLYsCyAspGlyThrArgIleAspGlnProCyAspProArg 1876  
4195 TGTCCCCCAATTACTTTGGGCACTATCTGTAGACACAGATTGACCAAGCTTGTCCCGCT 4254  
1877 GlyTrpTrpGlyHisProThrCyAspGlyProCyAsnCyAspValSerLYsGlyPheAsp 1896  
4255 GGTGTGTGGGAGCATCCCACTGTGGCCATGTGCAACTGTGATGTGACAAAGGCTTTGAC 4314  
1897 ProAspCyAsnLYsThrSerGlyGlyLYsCyAspLYsGlyAsnHisLYsArgProPro 1916  
4315 CCAGACTGCAACAAAGACAGCGGAGTGGCCACTGCAAGAGAAACCACTACCGGCCCCCA 4374  
1917 GlySerProThrCyAspLeuCyAspCyAspTrpProThrGlySerLeuSerArgValCyAs 1936  
4375 GGCAAGCCCACTGCTCTTGTGTGACTGTGACTCCCAAGGCTCTTGTCCAGAGTCTGT 4434  
1937 AspProGluAspGlyGlnCyAspProCyAspProGlyValIleGlyArgGlnCyAspArg 1956  
4435 GACCTTAGAGATGGCCAGTGTCCATGCAAGCCAGTGTCACTGGGCGTCACTGTGACCGC 4494  
1957 CyAspAsnProPheAlaGlyValThrThrAsnGlyCyAspGlyValAsnTrpAspSerCyAs 1976  
4495 TGTACAAACCTTTTGTGTGAGTGCACCAACATGGCTGTGAAGTAATGAACAGCTGC 4554  
1977 ProArgAlaIleGlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaAla 1996  
4555 CCAAGACGATTAAGGCTGGATGTGGTGGCCCTGACCCGCTTGGGCTGCTGTGCT 4614  
1997 AlaProCyAspProLYsGlySerPheGlyLYsThrAlaValArgHisCyAspGlyHisArgGly 2016  
4615 GCTCTCTGTCCCAAGGCTCTCTTGTGGAGCTGTGCGCCACTGTGATGACACAGGGGG 4674  
2017 TrpLeuProProAsnLeuPheAsnCyThrSerIleThrPheSerGlyLeuLYsGlyPhe 2036  
4675 TGGCTCCCCCAACCTCTTCACTGACGCTCACTCACTTCTGAACCTGAAGAGGGCTTC 4734  
2037 AlaGluArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAla 2056  
4735 GCTAGAGGCTTACAGCGGAATGACTCAGGCTTACGACAGGGGCTCCACAGACTAGCC 4794  
2057 LeuLeuLeuArgAsnAlaThrGlnHisIleThrAlaGlyLYsTrpPheGlySerAspValLYsVal 2076  
4795 CTGCTCTGTGGCAACGCGCACGAGCACACAGCTGGCTAATTCCGACGACGTCAGAGGTG 4854  
2077 AlaTrpGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeu 2096  
4855 GCTTACAGCTGGCAGCGGCTGTGGGCCCAAGAGGACCCAGCGGGGCTTTGGGCTG 4914  
2097 SerAlaThrGlnAspValHisPheThrGlyAsnLeuLeuArgValGlySerAlaLeuLeu 2116  
4915 TCTGCCACACAGACGTCGACTTCACTGAGAACTGTGCGGGGTGGGCAAGCCCTCTG 4974  
2117 AspThrAlaAsnLYsArgHisLYsTrpGlyLeuIleGlnGlnThrGlnGlyGlyLYsThrAlaTrp 2136  
4975 GACACAGCCAAACAGCGGCACTGGAGCTGATCAGACAGACAGAGGGTGGCACCGCTGG 5034  
2137 LeuLeuGlnHisLYsTrpGlyAlaLYsAlaSerAlaLeuAlaGlnAsnMetArgHisIleThr 2156  
5035 CTGCTTCAAGCATTAAGGCTTACGCGGCTTACGCGCTGTGCCCAAGACATGGGCACTTAC 5094  
2157 LeuSerProPheThrIleValIleThrProAsnIleValIleSerValValArgLeuAspLYs 2176  
5095 CTAAAGCCCTTTCACACATGTGCACGCGCCCAACTGTGATCTCCGTAAGTGGCTTGGACAA 5154  
2177 GlyAsnPheAlaGlyAlaLYsLeuProArgTrpGlyGlnAlaLeuArgGlyGlnGlnProPro 2196

Db	5155	GGGAACCTTGCGGGCCCAAGCTGCCCCGCTACAGAGCCCTGGGGAGACAGCCCCG	5214
Qy	2197	AspLeuGluThrThrValIleLeuProGluSerValPheArgGluThrProProValVal	2216
Db	5215	GACCTTGAGACAAGCTCATCTGGCTGAGCTGCTTCAGAGAGAGCCCCCGTGTCTC	5274
Qy	2217	ArgProAlaGlyProGluGluAlaGluGluProGluGluLeuAlaArgArgGluThrArg	2236
Db	5275	AGGCCCGAGAGCCCCGAGAGAGCCCCAGAGGCCAGAGAGCTGGACAGGCGACAGCG	5334
Qy	2237	HisProGluLeuSerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAla	2256
Db	5335	CACCCGAGCTGAGGCAAGGCGTGGCTGGCCAGGCTCAATCTACCCGACCTTGCC	5394
Qy	2257	GlyLeuLeuProHisAsnTyrAspProAspTyrAspSerLeuArgValProValArgPro	2276
Db	5395	GGGCTACGCTCATTAACCTATGACCCCTGACAGGCGACGCTTGAGAGCTCCCAAGCGCCG	5454
Qy	2277	IleIleAsnThrProValValSerIleSerValHisAspArgGluGluLeuLeuProArg	2296
Db	5455	ATCATCAACAACCCGCTGGTGAAGCATACGCTCCATATATATAGAGAGACTTCTCCCGG	5514
Qy	2297	AlaLeuAspTyrProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrLys	2316
Db	5515	GCCCTGACCAACCCGCTACAGGTGAGTTCCGCGCTGGAGACAGAGAGCGACCAAG	5574
Qy	2317	ProIleCysValPheThrPheAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAla	2336
Db	5575	CCCATCTGTGTCTTCTGGACCATTCATCTGGTCACTGGACAGAGGTGGTGTGGAGCC	5634
Qy	2337	ArgGlyCysGluValAlaPheArgAsnGluSerHisValSerCysGlnCysAsnHisGluMet	2356
Db	5635	AGAGGCTGTGAAGTGTCTTCGCGACATGAGAGCCACGTCACGCTCCAGTGCACACATG	5694
Qy	2357	ThrSerPheAlaValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeu	2376
Db	5695	ACGAGCTTCGCTGGCTCATGAGACCTTCTCGCGGAGAAATGGGAGATCTCTCCACTG	5754
Qy	2377	LysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhe	2396
Db	5755	AAGACACTGACATACGTGGCTCTAGAGTTCACCTTGCGCTTCTTGCTCACTTCTTC	5814
Qy	2397	PheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThr	2416
Db	5815	TTCCCTCACTCTTCGCTGATCTCGCGCTCCAAACCAACAGGCAATCCGACGTAACTGACA	5874
Qy	2417	AlaAlaLeuGlyLeuAlaGluLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuPro	2436
Db	5875	GCTGCGCTGGGGCTGGCTCAGCTGAGTCTTCCCTCGGGAATCAACAGGCTGACCTCCCT	5934
Qy	2437	PheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTyr	2456
Db	5935	TTTGCTGCACAGCTCATTTGCCATCTCTGGTGCACCTTCGTACCTTGACCTTTTCCGCG	5994
Qy	2457	AlaLeuLeuGluAlaAlaLeuHisGluTyrArgAlaLeuThrGluValArgAspValAsnThr	2476
Db	5995	GCTCTGCTGAGAGGCTTCGACCTTGACCGGGAATCACTATGAGGTGCGGAGATCAACACC	6054
Qy	2477	GlyProMetAspPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeu	2496
Db	6055	GGCCCAATCGCTTCACTACATACGTGGGCTGGGGCGTGGCTGCTTCATACAGAGGCTA	6114
Qy	2497	AlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyr	2516
Db	6115	GGCGTGGGGCTGGACCCGAGGAGCTTAACGGAAACCTGACTTGTGCTCTCCATCATAT	6174
Qy	2517	AspThrLeuIleThrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPhe	2536
Db	6175	GACAGCTCATCTGGAAGTTTCTGGCCCGGTGGGCTTTCGCGTTCATATAGTGTCTTC	6234
Qy	2537	IleuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluLysLys	2556

Dd	6235	CTGACATCTGGGGGCGCCGGGCTCCTGTGCTGCCAGGGGACAGGGCTTTGAGAGAAA	6294
Qy	2557	GIyProValIserGIyLeuGIInProSerPheAlaValIleuLeuLeuSerAlaThrTyr	2576
Dd	6295	GGTCTGTCTCGGGGCTGGACAGCCCTCTTGCGCGTCTCTGCTGTGAGGGCCAGCTGG	6354
Qy	2577	LeuLeuAlaLeuLeuSerValaIleuSerAlaPheThrLeuLeuPheIleuIleuPheAlaThr	2596
Dd	6355	CTGCTGGACAGCTCTCTGTGCACAGGACACCTCTCTTCCACTACCTCTTTGCTACC	6414
Qy	2597	CysAsnCysIleGIInGIyProPheIlePheLeuSerTyrAlaValLeuSerIlyGIuAlaI	2616
Dd	6415	TGCAATTGCATCCAGGGGCGCCCTTCATCTTCTCTCTCTATGTGGTCTTAGAGAGAGTCC	6474
Qy	2617	ArgIlyAlaIleuIlybLeuAlaCysSerArgIlybProSerProAlaIleuThrThr	2636
Dd	6475	CGAAGACACTCAAGCTGGCTGTGCAGCGGAGCGGACCGCCCTGACTGTGACACAC	6534
Qy	2637	LysSerThrLeuThrSerSerTyrAsnCybProSerProTyrAlaIleGIyArgLeuTyr	2656
Dd	6535	AAAGTCCACCTGACCTGTCTTACATGTGCCACGACCCCTTACGAGATGGCGGCTGAC	6594
Qy	2657	GIInProTyrGIyAspSerAlaGIySerLeuHisSerThrSerArgSerGIyLySerGIIn	2676
Dd	6595	CAGCCCTTACGGAGACTGGCGGCTCTCTGCACAGCACAGTCCGCTGGGGCAGAGTCAAG	6654
Qy	2677	ProSerTyrIleProPheLeuLeuArgGIInIleuSerAlaIleuAsnProGIyGIInGIyPro	2696
Dd	6655	CCGACGTACACATCCCTCTTGTGTAGAGAGAGAGTCCGACATGAACCTGTGGCAGAGGGCC	6714
Qy	2697	ProGIyLeuGIyAspProGIySerLeuPheLeuGIInGIyGIInAspGIInHisAspPro	2716
Dd	6715	CCTGGCCTGGGGGATCCAGGAGCGCTGTCTCTGTGAAGGTCAAGACCGACAGCATGATCT	6774
Qy	2717	AspThrAspSerAspSerAspLeuSerLeuGIInAspAspGIInSerGIySerTyrAlaIser	2736
Dd	6775	GACACGGACCTCCAGACATGACCTGTCTTGTAGAAACACACAGATGGCTCTTAGTCTCT	6834
Qy	2737	ThrHisSerSerAspSerGIInGIInGIInGIInGIInGIInGIInGIInAlaIlePhePro	2756
Dd	6835	ACCACATCATAGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCGCCCTTCCCT	6894
Qy	2757	GIyGIInGIInGIyTPAspSerLeuLeuGIyProGIyAlaGIuAlaIleuProLeuHisSer	2776
Dd	6895	GGAGAGACAGGGCTGTGGATGCTGTCTGGGCTGTGACACAGAGACATGGCCCTTGCACAGT	6954
Qy	2777	ThrProIlyAspGIyGIyProGIyLyAspAlaProTyrProGIyAspPheGIyThr	2796
Dd	6955	ACTCCCAAGATGGGGGCCCAAGGGCTGTGGCAGGCCCTGTGGCCAGAGACTTTGGGACC	7014
Qy	2797	ThrAlaIlybGIInSerSerGIyAsnGIyAlaProGIInIuArgLeuArgGIuAsnGIyAsp	2816
Dd	7015	ACACAAAGAGATTAATGTGGCAGAGGGGGCCCTAGAGAGCGGCTGGGAGAAATGAGAT	7074
Qy	2817	AlaIleuSerArgGIInGIySerLeuGIyProIleuProGIySerSerAlaGIInProHisIlys	2836
Dd	7075	GGCCGTGTCAAGAGGGGTCTCTAGGGCCCTTCCAGGCTCTTGTGCCACCTTCCACAA	7134
Qy	2837	GIyIleuLeuIlybIlybCybLeuProThrIleSerGIuIlySerSerSerLeuLeuArgLeu	2856
Dd	7135	GGCATCTTAAAGAAAGATGTCTGCCACCATCAAGCGAGAAAGACAGCTTCTCGGGCTC	7194
Qy	2857	ProLeuGIInGIyThrGIySerSerArgIySerSerAlaSerGIInGIySerArgIy	2876
Dd	7195	CCCTCGAGCATCAACAGGGGTCTTCCGGGGCTCTCCCTATGTAGAGGACGGGGGC	7254
Qy	2877	GIyProProAlaGIyProProProArgGIInSerLeuGIInGIInLeuAsnGIyValMet	2896
Dd	7255	GGCCCCCTCCCGGCCACCGCCCGCGAGAGCTTCCAGAGAGAGCTGAAGCGGGGTCAATG	7314
Qy	2897	ProIleAlaMetSerIleIyAlaGIyThrValAspGIuAspSerSerGIySerGIuPhe	2916
Dd	7315	CCCATGTGCATAGACATCAAGCAGGACCGGTGATGAGACTGTCAAGGCTCCGAAATT	7374

Qy	291	Leuphosphaphenleus 2923
Dn	7375	CCTCCTTAACTTCCTGCAT 7395
RESULT 8		
ID	AAS06332	standard; cDNA; 9045 BP.
XX	AAS06332:	
AC	26-SEP-2001	(first entry)
DT		
DE	DNA encoding seven-pass transmembrane receptor-like protein, MEM1.	
XX		
KM	Seven-pass transmembrane receptor; MEM1; therapeutic; diagnostic; MEM2;	
KM	human; Alzheimer's disease; Parkinson's Disease; cancer; nephrology;	
KM	female reproductive health; lung disorder; brain disorder; schizophrenia;	
KW	heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;	
KM	cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;	
KM	vision-related disorder; neoplastic pathology; MEM7; MEM8; ss.	
OS	Homo sapiens.	
FH		
FT	Key	Location/Qualifiers
FT	CDS	1..9045
FT		/tag= a
FT		/product= "Seven-pass transmembrane receptor-like
PN		protein, MEM1"
XX		
PD	WO200144473-A2.	
PR	21-JUN-2001.	
PF	14-DEC-2000; 2000MO-US33909.	
PR	14-DEC-1999; 99US-0170564.	
PR	27-DEC-1999; 99US-0173165.	
PR	27-DEC-1999; 99US-0173162.	
PR	29-DEC-1999; 99US-0173544.	
PR	04-JAN-2000; 2000US-9966564.	
PR	09-AUG-2000; 2000US-0223529.	
PR	13-DEC-2000; 2000US-9966565.	
XX		
XX	(CURA-) CURAGEN CORP.	
PI	Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spyrek KA;	
DR	WPI; 2001-398154/42.	
DR	P-PsDB; AAU02196.	
XX		
PT	Novel polypeptide comprising members of protein families (e.g.,	
PT	seven-pass transmembrane receptor proteins) according to presence of	
PT	domains and sequence relatedness are useful for treating or preventing,	
PT	e.g., Alzheimer's and Parkinson's -	
XX		
PS	Claim 9; Fig 1; 162pp; English.	
XX		
CC	The sequence represents the coding sequence of seven-pass transmembrane	
CC	receptor-like protein, MEM1, selected from a group (MEM1-MEM8) comprising	
CC	members of protein families according to the presence of domains and	
CC	sequence relatedness, e.g., seven-pass transmembrane receptor protein	
CC	(MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),	
CC	phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).	
CC	The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are	
CC	all useful for treating or preventing a pathology associated with (I)	
CC	comprising administering (I), (II), or (III) to a subject (preferably a	
CC	human). In addition, (I), (II), and (III) may be used to manufacture a	
CC	medicament for treating a syndrome associated with a human disease that	
CC	is associated with (I). Furthermore, (I) may be used to identify agents	
CC	that bind to it, screen modulators of its activity and determine the	
CC	presence or predisposition to a disease associated with altered levels of	
CC	(I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease.	

[illegible]

```

Db      724 AGCCGTAAGTTCCGAGTCGCACTACGAGTGGCGTTTGGAGAACGACCGGCGGC 783
Qy      198 ThrProValAlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGlu 217
Db      784 ACCCTCATCTCCAGCTGACGCGCACTACCATGACGGGCGAGGAGGAGCGCGTGGAC 843
Qy      218 TyrThrMetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProVal 237
Db      844 TATTATCATGAGAGGGCGTGTTCGACGAGCGCTCCGGGGGCACTTCGGAATTCAGCTGTGC 903
Qy      238 ThrGlyAlaValThrThrAlaGluGluLeuAspArgGluThrIleSerThrIleValPhe 257
Db      904 ACCGGGCGCGTGGACGACGACGCTGACCGCGGAGCCAGAACGACGACGACGCTCC 963
Qy      258 ArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle 277
Db      964 AGGGTAAGCCCTGAGCTACGACTACGCGCGCGCTCCGACCACTACCTACCTGTC 1023
Qy      278 LeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGluThrIleArgLeuSer 297
Db      1024 TTGCTCAAGACACCAACGACCAAGCCCGCTTCGAGAGTCCGAGTACCGCGAGCGC 1083
Qy      298 LeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAsp 317
Db      1084 GTGCGGAGAACCTGAGGTGGCTACGAGTGTCTACCATCCGCGCACGACCGCGAC 1143
Qy      318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyIleSerProSer 337
Db      1144 TCGCCATCAACGCCAATCTGCTTACCGGTGTGGGGCGCGCG----- 1191
Qy      338 GluValPheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArg 357
Db      1192 GACGCTTCACGCTCAACGAGAGCTGTGGGTGTGAGACACGCGCGGTGTGACCGCG 1251
Qy      358 GluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGly 377
Db      1252 GAGGAGCGCGCCAGTACGAGCTCTGTGTGAGGCGACGACGAGGCGGCAATCCGCGC 1311
Qy      378 ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaPro 397
Db      1312 CCCTCATGTGCCAGCGCACCGCTGTATCATCGAGTGTGAGAGCAAGAACCACTACCCC 1371
Qy      398 GlnPheSerGlyIleArgTyrValValGlnValArgGluAspValThrProGlyAlaPro 417
Db      1372 CAGTTCAAGCAGACAACTACGTGTCCAGGTCCCGAGGACGTGGGGCTCAACACGGCT 1431
Qy      418 ValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValValHisTyrSer 437
Db      1432 GTGCTCGAGTGCAGGCGCACGAGCCGCGACGAGGCGACGAGCCGCACTTCATCACGC 1491
Qy      438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp 457
Db      1492 ATCTCAGCGGGAACGTGGCGCGCGCACTTCTACCTGACCTGAGCGGAGTCTGGAT 1551
Qy      458 ValValSerProLeuAspTyrGluThrThrIleGlySerGlyThrLeuArgValArgAlaGln 477
Db      1552 GTGATCAACCCCTGGATTTCGAGATGTCGAAGAAATCTCGCTGACATTAAGCCGACG 1611
Qy      478 ArgGlyValArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497
Db      1612 GAGTGGGGCGCGCGCGCTCATCTTCAGGGGTGTGTCTGTGACGAGTCTGGAT 1671
Qy      498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer 517
Db      1672 GTCAAGACACACAGCTATCTTTGTGAGCAGCCCTTCAGGCGCACGGTGTGAGGAAAT 1731
Qy      518 ValProLeuGlyTyrLeuValIleuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537
Db      1732 GTGCCCTTGCGGTACCCCGTGTGCAATTCAGGCGGTGACCGGAGCTCTGAGAGAAAC 1791
Qy      538 AlaArgLeuGluIleValArgLeuAla-----GlyValGlyHis----- 549

```

```

Db      1792 GCCCGGCTGACATACGCTGTGTGACACGGGCTTCACCTTTCTGGGGGGCGGACGGCT 1851
Qy      550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1852 GGGCTTAAGAAATCTCGGCCCAACCCCTGACTTCCCTTCGAGATTCACAAACGCTCGGT 1911
Qy      561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      1912 TGATCATCAGTGTGTGCCGAGCTGAGCGGAGGAGGTGAGACACTACGCTTCGGGGTGT 1971
Qy      581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      1972 GAGGCGGTGACCAACGGCTCCCGCCCAATGAGCTCTCCACACGCTGTTCATCAGGTG 2031
Qy      601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      2032 CTGAGCGTAATGACACAGACCCGGGTTCACGAGCCCACTACGAGCTTCGTGAAAT 2091
Qy      621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      2092 GAGGATCGCGCGGTGGGAGACGCGTGTGACCTTCGACGCGCGCGACCGTGAAGCCAAC 2151
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      2152 AGTGTATTACTTACCACTCACTCAAGCGCGCAACCCGGAACCGCTTGTGACTCAAGCAGC 2211
Qy      661 GlnSerGlyGlyGlyLeuValSerIleLeuAspProLeuAspTyrTyrLeuGluArgGln 680
Db      2212 CAGAGAGGGGGCGGCTCATACCTTCGCGCTTACCTCGACTACCAAGACAGACAGAG 2271
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      2272 TACGTGTGGCGGTGACAGCATCCGACGCGCACAGGTGCGACACTGCGCATGCTTAATC 2331
Qy      701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      2332 AACGTCACTGAATGCCAACACCAAGCGCTGTCTTTCAGACTCCCATTAACAAGTAGAT 2391
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db      2392 GTCACTGAGACAGGCGCTGTGGGCACTTCGTAACCTCAAGTCCACAGATAGAGAC 2451
Qy      741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2452 ACAGGAGGAATGCCCGCATCACTACGTGATTCAAGACCCCGTGGCCAGATTCCGCAAT 2511
Qy      761 AspAlaAspThrArgValAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlyValSer 780
Db      2512 GACCCCGACAGTGGACCATATACATGACATGAGTGGAGCTGACTATGAGAACCAAGTCCGC 2571
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db      2572 TACAGCTGACCATATATAGCCCAAGAGACAAAGGATCCCGGAAATCAGACCAACCAAC 2631
Qy      801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgSerTyr 820
Db      2632 CTAGAGATCCCATCTTCGATGCGCAATGACATGACACCCAGTTCCTGTGGAAATTTCAAC 2691
Qy      821 GlnGlySerValTyrGluAspValProProPheThrIleSerValLeuGlnIleSerAlaThr 840
Db      2692 CAGGCTCATCTTGAAGAGTCTCACTTCGACACGACATCTTCAGGTCTTCGACAG 2751
Qy      841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly 860
Db      2752 GACCGGAGCTCAGGCTCCCAATGGGGGTGTGTGTACCTTCAGAGGTGGGAGACGACGGC 2811
Qy      861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db      2812 GATGGGAGCTTCTCATCTAGACCCACGTCGCTGATGTCCACCAAGCGCGGTGGAC 2871
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetPro--- 899
Db      2872 CCGAGGAATGTGGCGGTATCAACCTTTGGGCTGTGGTGTGATCGGGGACATCCCACT 2931

```

QY	900	ProAlaArgThrPrometGluValThrValThrValIleuAspValAsnAspAsnProPro	919
Db	2932	CCCCTTACCGCCTCGGTAGAAATCCAGAGTACCATCTTGAGCAATTATGACAAATGCCCC	2991
QY	920	ValPheGluGlnAspGluPheAspValPheValGluGlnAsnSerProIleGlyIleuAla	939
Db	2992	ATGTTTGAAGAAGAACTGAGGCTGTTTGTGTGAGAAACAAACCCAGTGGGTCGGTGTG	3051
QY	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrosGln	959
Db	3052	GTCGCAAAAGATTCCGTACTAAGACCCCTATATGAAGCCCTAATGCCAATCATATATACG	3111
QY	960	IleValGluGlyAsnIleProGluValPheGluIleuAspIlePheSerGlyGluIleuThr	979
Db	3112	ATTGTGGAAAGGGGACATCCGCAATTTCTTCCAGCTGGACCTGCTCAAGGGGACCTGGCT	3172
QY	980	AlaIleuValIaAspIleuAspTyrosGluAspArgProGluTyrosValIleuValIleGlnAlaThr	999
Db	3172	GCCATGTGTGAGACTGTGACTTTGATGCTCGGCGGAGTATGTGCTGTGTGTGTGAGCCACG	3231
QY	1000	SerAlaProIleuValSerArgAlaThrValHisValArgIleuAspArgAsnAspAsn	1019
Db	3232	TCCGCTCCGCTGTGTGAGCCGAGCCACCGGTGACATCTTCTGTGTGACACAGATATACAC	3291
QY	1020	ProProValIleuGlyAsnPheGluIleuPheAsnAsnTyrosValThrAsnArgSerSer	1039
Db	3292	CCGCTGTGTGCGCCGACCTTCAGATCTCTTCAACAATATGTCAACAAACAGTCCAC	3351
QY	1040	SerPheProGluGlyAlaIleGlyValArgValProAlaHisAspProAspIleSerAspSer	1059
Db	3352	AGTTTCCCCACCGCGGTATGTGGCTGTGCATCCCGGCCATATACCCGAGTGTGCACACG	3411
QY	1060	LeuThrTyrosPheGluArgGlyAsnGluIleuSerIleuValIleuIleuAsnAlaSerThr	1079
Db	3412	CTCAACTACACCTTGTGTGACAGGGCAACAGAGTGGCGCTGTGTGCTGTGACCCCGCACG	3471
QY	1080	GlyGluIleuTyrosLeuSerArgAlaIleuAspAsnAsnArgProIleuGlnAlaIleuSer	1099
Db	3472	GCGCAACTGCAAGCTCAGCCGCGACCTTGACAAACACCGCGCTGGAGCGCTCAATGAG	3531
QY	1100	ValIleuValSerAspGlyValHisSerValThrIaGlnCysAlaIleuArgValThrIle	1119
Db	3532	GTTGTGTGTGTGATGAGCATTCACAGCTTCACGCGCTTCTGCACTTCGCTGTACCATC	3591
QY	1120	IleThrAspGluMetLeuThrHisSerIleThrIleuArgIleuAspMetSerProGlu	1139
Db	3592	ATCACGCGACATGTGTGACCAACAGCATCTGTCGCGCTGAGAAATGTCCACGAG	3651
QY	1140	ArgPheIleuSerProIleuGluGlyIleuPheIleGlnAlaValAlaIleThrIleuAlaThr	1159
Db	3652	AAAGTTCCTGTCCCGCTGTGCGCTCTTTCGTGAGGGGGTGGCGCGCTGTCTCCAC	3711
QY	1160	ProProAspHisValValIlePheAsnValGlnArgAspThrAspAlaProGlyGlyHis	1179
Db	3712	ACCAAGACGACGTCTTGTGTCTTCAACTCCAGAACGACACCGACGTC---AGCTCCAAC	3766
QY	1180	IleIleuAsnValSerIleuSerValGlyGlnProProGlyProGlyGlyProProPhe	1199
Db	3769	ATCTCGAAGTGACCTTCTCGCGCGCTGTGCGCTGTGCGCGCGCTCCCGGCGC-----CAGTTC	3822
QY	1200	LeuProSerGlnAspIleuGlnGluArgIleTyrosAsnArgSerIleuIleuThrAlaIle	1219
Db	3823	TTCCTGTGTGAGACCTCAGAGACGATCTACTGAATTCGAGCGCTGTCTACCACTCC	3882
QY	1220	SerAlaGlnArgValIleuProPheAspAsnIleCysIleuArgGluProCysGluAsn	1239
Db	3883	TCCACGCGACGCGTCTGCTCCCTTCGACGACAAACATCTGCTGGCGGAGCCCTGGAGAAC	3942
QY	1240	TyrosIleArgCysValSerValIleuArgPheAspSerSerAlaProPheIleAlaSerSer	1259
Db	3943	TACATGAAGTGCAGTCTCGCTTCGTGGAATTGACACAGCTCCGCGCCCTTCTCAGCTCCAC	4002

QY	1260	SerValleuPheuaqProIlehiProValGelyleuAqCyBaGyCyProProGelY	12719
Db	4003	ACCGGCTCTTCGGCCCATCAACCCCAACAAGCCCTGCCTGGCCGGCCGGC	40628
QY	1280	PheThrngIyaAPYrCYeGluWthGluValaLapleuCyYrSerArgProCyegIyPro	12999
Db	4063	TTCAACGGCGCATCTAGTGGAACCGAATACCACTCTGTCTACTCGAACCCGTGGCGCC	41222
QY	1300	HiGelYaqCyBaGSerArgGluGelyleYrThrCyBleuCyBaGApGelYrThr	13155
Db	4123	AACGGCCCTCTCCGACGCGGAGGGCGGCTACCACTGCGAGTGGCTTGAGAGACTTCACT	41872
QY	1320	GlyGluHsGySgIuValaSerLAPaGSerGelYArgCySthrProGelYValCYuByspN	13332
Db	4183	GGAAGACACTGTAGAGGTGATGCCCGGCTCAAGCGCTGTGCCAAGGGGTGTGCAAGAC	42424
QY	1340	GlyGelYrThCyVaLaSmLeuLeuValGelylePheLysCyBaAPCyProSergIyAap	13559
Db	4243	GGGGGCACTGCGTAACCTGCTCATGCGGGCGGCTTCCATGCGTGTGCTCTGGCGAG	43020
QY	1360	PheGluYrProYrCYeGlnValIThrArgSerPheProLahIaSerPheIleThr	13799
Db	4303	TATAGAGGCTTACTGTAGGTACCAACGAGGCTTCCGGCCCACTCTTGATCAAC	43622
QY	1380	PheArgGlyLeuAArgGlnArgPhehiSphethrLeuAlaLeuSerPheAlaThrGelY	13999
Db	4363	TTCCGGGGCTCGAAGACAGCGCTTCACTTCAACATCTCCCTCAGGTTGCCATCAGAA	44222
QY	1400	ArgApGelyleuLeuLeuYrArgnGlyArgPheanGluYrShIaAPPheValAlaLeu	14119
Db	4423	AGGAACGGCTTGGCTTCTTCAACAGCGCGCTTCATAGAGAACGACGACTTCATGCGCTG	44822
QY	1420	GluValIleGlnGluGlnValGlnLeuThrPheSerLahGelYusSerThrThrThrVal	14319
Db	4483	GAGATCGTGAAGACAGACAGTGCACCTCTCTGTGAGCGAGACAAACAACGCTG	45242
QY	1440	SerProPheValProGelYGelYValaSerApGelYnIThrPhIsthrValGlnLeuYrYr	14559
Db	4543	GCAACGAAAGGTTCCAGTGGTGTGAGTGAACGAGCGGTGGCACTGTGTCAAGTGCAGTAC	46022
QY	1460	ThrThnLysProLeuLeuGelYnIThrGlyLeuProGlnGelYProSergIuGlnLysVal	14719
Db	4603	TACACAAAGCCCAATATTGTGGCACTGGCGCTTGCCCAATGGCGGTCCGGGAAAAAGTGG	46622
QY	1480	AlaValValThrValaAPGelYCyAAPThrGelYValAlaLeuAArgPheGelYSerValLeu	14919
Db	4663	GCGGTGGTGAAGTGAATGTGTGACACAAACATGCTGTGGCCCTTTGGAAAAAGACATC	47222
QY	1500	GlyAntYrSerCyBaLahGelngIYThrGlnGelYGelYSerLysSerLeuAapLeu	15119
Db	4723	GGGAACTAAGCTCGCTGCCGACGGGCACTCAAGCCGGCTTCAAGAGTCCCTGTGATGTG	47822
QY	1520	ThrGlyProLeuLeuLeuGelYValProAapLeuProGiuserPheProValArgMet	15319
Db	4783	ACCGGCTCTACTCTCGGGGGGTGTCCCAACTGCGCAAGACCTTCCAGTGCACAAAC	48424
QY	1540	ArgGlnPheValGelYsMeLArgAmlLeuGlnValaAPSerArghiSilaAPMeLala	15559
Db	4843	CGGCAAGTTCGGGCTCATATCGGAACCTGTCACTGACCGGCAAAATGTGCATGCGCC	49020
QY	1560	AspPheIleAlaAanAanGelYThrValProGelYCyAProAlaLysLysValaCyAap	15719
Db	4903	GGATTTCATCGCAACATGGCAACCGGAAAGGCTGCGCTGTGGAAGAACTTCTGAGT	49622
QY	1580	SerAnthrCYuHlaaGelngIYThrCYuAlaAanGlnIThrAPaLahIaSerCYeGluY	15919
Db	4963	GGAGAGCGGATGCAATAGAGAGGCACTGTGTCAACAGGTGAATATGTATCTGTGAG	50222
QY	1600	CyAProLeuGelYrPheGelyleYrYrSerCYuAlaGlnLeuMeLaaanProGlnhiSph	16119
Db	5023	TGTCACTCGCAATTCGGCGGAAAGAACTGTGAGAACCAATGCTCAACCCCACTCTTC	50822
QY	1620	LeuLysSerSerLeuValaLTPHIGlyLeuSerLeuProLieserGlnProTTPYr	16319







OS	Homo sapiens.
XX	
PN	NC0200175093-A1.
PD	11-OCT-2001.
PF	30-MAR-2001; 2001WO-US10484.
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PR	22-SEP-2000; 2000US-0686860.
PR	23-OCT-2000; 2000US-0695618.
PR	30-NOV-2000; 2000US-0728711.
PR	14-MAR-2001; 2000US-0728711.
PA	(HYSEQ-) HYSEQ INC.
PI	Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Xu C,
PI	Yang Y, Zaho QA, Chen R, Wang D, Goodrich RW, Liu C, Drmanac RT;
XX	
DR	WPI; 2001-626432/72.
XX	P-PDSB; AAU68533.
PT	New polypeptides and nucleic acids, useful for diagnosis, treatment of
PT	inflammatory, autoimmune, neurological, myeloid or lymphoid cell, bone
PS	dysgenerative disorder, cancer and promoting wound healing -
XX	
XX	Claim 1; Page 164-177; 336pp; English.
CC	The invention relates to isolated human polypeptides (which may be
CC	cytokines) and the polynucleotides encoding them. The protein is useful
CC	for identifying a compound which binds to it (e.g. modulators, agonists
CC	and antagonists). The polynucleotides are useful as an array for mismatch
CC	detection. The proteins and nucleic acids are useful as nutritional
CC	sources or supplements. The protein exhibits activity relating
CC	to cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC	stem cell growth factor activity, immune stimulating or immune
CC	suppressing and activin or inhibin related activities. The proteins (and
CC	antibodies raised against them) and nucleic acids are therefore useful in
CC	the diagnosis and treatment of diseases and disorders such as cancer,
CC	central and peripheral nervous system diseases and neuropathies,
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, spinal cord disorders, head trauma, cerebrovascular
CC	diseases, stroke, myeloid or lymphoid cell disorders, platelet disorders
CC	thrombocytopaenia, stem cell disorder, aplastic anaemia, for
CC	regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC	growth, and in tissue repair, healing of burns, incisions, ulcers, for
CC	treating osteoporosis, osteoarthritis, bone degenerative injuries, or
CC	periodontal disease, lung or liver fibrosis, reperfusion injury in
CC	various tissues, various immune deficiencies and disorders including
CC	severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC	autoimmune disorders (e.g. multiple sclerosis, rheumatoid arthritis,
CC	diabetes mellitus, myasthenia gravis), allergic reactions and conditions,
CC	such as asthma or other respiratory problems, coagulation disorders,
CC	haemophilia, septic shock, sepsis, arthritis, nephritis and inflammatory
CC	bowel disease, viral infection and are useful in altering bodily
CC	characteristics. The present sequence encodes a novel protein of the
XX	invention.
SQ	Sequence 11389 BP; 2217 A; 3596 C; 3401 G; 2175 T; 0 other;
Alignment Scores:	
Pred. No.:	0 Length: 11389
Score:	8781.50 Matches: 1719
Percent Similarity:	69.42% Conservative: 422
Percent Local Similarity:	55.74% Mismatches: 689

[illegible]



Dh 3112 ATTGTGAGAGGAGCATGCGGCAATTTCTTCAGCTGAGCCTGCTCAACGGGGAGCTGCGT 3171  
Qy AAlaLeuValAspLeuAspArgProGluValLeuValIleGlnAlaThr 999  
Dh 3172 GCCATGTGAGGTGAGCTTGTAGGCTCCGGGGAGTATGTGCTGTGTGTCAGGGGCAAG 3231  
Qy 1000 SerAlaProLeuValSerAlaThrValIleValArgLeuAspArgAsnAspAsn 1019  
Dh 3232 TCGGCTCCGTGTGAGGCGAGCCAGGCTGACATCTCTCGTGGACCAATGACAAAC 3291  
Qy 1020 ProProValLeuAlaLeuPheGluIleLeuPheAsnAspThrValThrAsnArgSer 1039  
Dh 3292 CCGCTGTGTGCGCCAGCTTCAGATCTCTTCAACAACATATGTCACCAACAGTCCAC 3351  
Qy 1040 SerPheProGluValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Dh 3352 AGTTTCCCAACCGGCGGATCGGCTGATCCCGGCCATGACCCCAAGTGTGACAGC 3411  
Qy 1060 LeuThrTyrSerPheGluArgGluValAsnGluLeuSerValLeuLeuAsnAlaSerThr 1079  
Dh 3412 CTCACACTACACTTCGTCGAGGGCAACGAGCTGCGCTGTGCTGAGACCCCGGCAAG 3471  
Qy 1080 GAGluLeuValLeuSerAlaAlaLeuAspAsnAspArgProLeuGluAlaIleMetSer 1099  
Dh 3472 GGGAGACTGAGCTCAGCCCGGACCTGGACAAACACCGCGCTGAGGGGCTCATAGAG 3531  
Qy 1100 ValLeuValSerAspGluValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Dh 3532 GTGTCTGTCTATGTGATCCACAGGCTCAGCGGCTTTCGACCCCTGCTGTACATC 3591  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Dh 3592 ATCAGCGACACATGTGACCAACAGCATCATCTCGCGCTGGAGAACATGTCCACAGAG 3651  
Qy 1140 ArgPheLeuSerProLeuLeuGluValPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
Dh 3652 AAGTCTCTGTCGCGCTGCGCTCTTCTGTGAGGGGGGAGGCGCGCTGTCTCACC 3711  
Qy 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGluGlyHis 1179  
Dh 3712 ACCAAGACACAGCGTCTGTCTTCAACGTCAGAACACACCGACGCTC---AGCTCAAC 3768  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGluProGluGlyProProPhe 1199  
Dh 3769 ATCTTGAACTGACCTTCTCGGCGCTGCTGCGCGCTGCGCGCTC---CAAGTTC 3822  
Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuThrAlaIle 1219  
Dh 3823 TTCCTGTGAGAGACCTGACAGACAGACATCTAATGATGAGAGCTGTGACCAACATC 3882  
Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Dh 3883 TCACAGCAGCGCGTGTGCTGCTTCAACGACATCTGCGTGGCGAGCCCTGGAGAAC 3942  
Qy 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Dh 3943 TACATGAAAGTGGCGTCTGTCGATTCGACAGCTCCGCGCTTCTCAAGTCCAC 4002  
Qy 1260 SerValLeuPheArgProIleHisProValGlyLeuLeuArgCysArgCysProProGly 1279  
Dh 4003 ACCGTGCTCTTCGCGCCATCCACCCATCAACGCGCTGCGCTGCGCTGCGCGCGC 4062  
Qy 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Dh 4063 TTCACCGGCGACTCTGAGAGACGGAAGATGACCTTGTACTCCGACCGGTGGGCGCC 4122  
Qy 1300 HisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Dh 4123 AACGCGCGCTGCGGAGCGGAGCGGCTACACCTGCGAGTCTTCGAGACCTTCACT 4182  
Qy 1320 GlyIleHisCysGluValSerAlaArgSerGlyArgCysThrProGluValCysLeuAsn 1339  
Dh 4183 GGAAGACACTGTGAGGTGAGTCCGCTCAGCGCGCTGTCCAAAGGGGTGTGGAAGAC 4242

Qy 1340 GlyIleThrCysValAsnLeuValGlyGlyPheLysCysAspCysProSerGlyAsp 1359  
Dh 4243 GGGGACACTGTGTACCTGTCTATCCGGCGGCTTCACTGTGTGTCTCTCTGCGAG 4302  
Qy 1360 PheGluValProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Dh 4303 TATGAGAGGCGCTTACTGTGAGGTGACCAACAGAGACTTCCCGCCAGTCTTCTCACC 4362  
Qy 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrArgGlu 1399  
Dh 4363 TTCGGGCGCTGAGACAGGCGCTTCACTCAACATCTCCCTCAGCTTGTGCAAGAA 4422  
Qy 1400 ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluValHisAspPheValAlaLeu 1419  
Dh 4423 AGGAACGGCTTGTCTCTCAACAGCGCGCTTCAATGAGACAGCATTCATCGCCCTG 4482  
Qy 1420 GluValIleGlnGluGlnValGluLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
Dh 4483 GAGATCTGTGACAGACAGGTGACCTCACTTCTGTGAGGCGAGAACAAACGACCGTG 4542  
Qy 1440 SerProPheValProGluValSerAspGluGlnThrHisThrValGluLeuValTyr 1459  
Dh 4543 GACCGAAGTTCCTCACTGTGTGTGAGTACGCGCGGTGACACTGTGTGACAGTGCATTC 4602  
Qy 1460 TyrAsnLysProLeuLeuGluGlyGlnThrGlyLeuProGlnGlyProSerGluGlnVal 1479  
Dh 4603 TACAACAAGCCCAATATGTGGCCACTGGGCGCTGCCCATGGGCGCTCCGGGGAAAGATG 4662  
Qy 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaAlaLeuArgPheGlySerValLeu 1499  
Dh 4663 GCGGTGTGACAGTGAATGATGTGACAAACAGTGGCTGTGTGGGAAAGACATC 4722  
Qy 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLeuSerLeuAspLeu 1519  
Dh 4723 GGAATCAAGCTGTGCGCTGCCAGGCACTGACCGGCTCCAAAGTCCCTGTGATCG 4782  
Qy 1520 ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet 1539  
Dh 4783 ACCGCGCTCTACTCTGTGGGGGTGTCCCAACCTGGCAAGAGCTTCCACATGCAAC 4842  
Qy 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
Dh 4843 CGCGTGTGTGGGCTGACATGCGAACCCTGTGATGACAGGCAAAATGTGACATGGCC 4902  
Qy 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLeuValAsnValCysAsp 1579  
Dh 4903 GGAATCATGCCAACAATGGACCCGGGAAGCTGTGCGCTCGGAGAACTTTCGGAT 4962  
Qy 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTTPAspAlaPheSerCysGlu 1599  
Dh 4963 GGGAGCGGTGTCAGAATGAGGACCTGTGTCAACAGGTGAAATATGATGTGTGTAG 5022  
Qy 1600 CysProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
Dh 5023 TGTCACTCCGATGTGGCGGAAAGACTGTGAGCAAGCCATCTCAACCCCAAGCTTTC 5082  
Qy 1620 LeuGlySerSerLeuValAlaTTPHisGlyLeuSerLeuProIleSerGlnProTyr 1639  
Dh 5083 AGCGGTGAGAGGTGTGTCTGAGTACCTGAAGTACATCATCTGTGCTGTGCTGATC 5142  
Qy 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
Dh 5143 CTGGGCTCATGTTCGAGACCGGAAAGGAGCAGGGTCTATATGAGGCAACATGTGT 5202  
Qy 1660 GlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGly 1679  
Dh 5203 GGGCCACCAAGCTTGTGCTTCAAGATCTGGAACACTACTCTCAAGTTTGAAGTCCAC 5262  
Qy 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTTP 1699  
Dh 5263 GGGCCCTCGAATGTGAGTCCGTGATGTCTGCGGGTTCGGGTGACCGAGCGGAGTGG 5322





Db 7483 ATGCTGCGCTCCACCTGCACAGCATTACAGACACTCGCGTGGCGCTTCTCTCTCT  
 Qy 2423 GlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIle 2442  
 Db 7543 CAGCTGGTGTTCGATGGATGATACACAGACGAAACCGTTTCTGTGACAGAGGT 7602  
 Qy 2443 AlaIleLeuLeuHisPheLeuTyLeuCysThrPheSerTrpAlaLeuLeuGlnAlaLeu 2462  
 Db 7603 GCATCCCTCCCTCACTACATCATGAGACACTTTCCTGACACCTCGAGAGAGCTG 7662  
 Qy 2463 HisLeuTyTrpAlaLeuThrGlnValArgAspValAsnThrGlyProMetAlaPheTy 2482  
 Db 7663 CATGTACCGCAGTCGACCGAGAGTGGCAGACATCAACAGCGGCGCCATCGGTTCTAC 7722  
 Qy 2483 TyMetLeuGlyTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspPro 2502  
 Db 7723 TAGTGTGGGTGGGGCATCCGGCATTCATGACAGACTGGCGGTGGCTGGACCC 7782  
 Qy 2503 GlnGlyTyTrpGlyAsnProAspPheCysTrpLeuSerIleTyTrpAspThrLeuIleTrpSer 2522  
 Db 7783 CAGGGCTACCGGAAACCCGCACTTCGTGGCTGCTGCAAGACACCTGATTGGAGC 7842  
 Qy 2523 PheAlaGlyProAlaAlaPheAlaValSerMetSerValPheLeuTyTrpIleLeuAlaIle 2542  
 Db 7843 TTTCGGGGGCGCCATCGAGCTGTTATATCATCAACAGTCACTTCTGTCTATGTGCA 7902  
 Qy 2543 ArgAlaSerCysAlaIleAlaGlnArgGlnGlyPheGlnIleValSerGlyProValSerGlyLeu 2562  
 Db 7903 AAGTTTCCGCGCAAGAAAGAACCATATTATGGAGAAAGAGATCGTCTCCGCTG 7962  
 Qy 2563 GlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSer 2582  
 Db 7963 AGACCGCATCTCTCGCTGCTGCTCATCAGCGCCACTCGGCTGGGGCTGCTGGCT 8022  
 Qy 2583 ValAsnSerAspThrLeuLeuPheHisTyTrpLeuPheAlaThrCysAsnGlyIleGlnGly 2602  
 Db 8023 GTAAACCGCATGATGATGACCTTCACTACCTCTGCGCATCTTCAGCGCTTACAGGGC 8082  
 Qy 2603 ProPheIlePheLeuSerTyValValLeuSerIleGlnValArgIleAlaLeuIle 2621  
 Db 8083 CCTTCGTCCTCTTTTCCACTGCGTGTCAACAGAGAGTCCGGAAGACACCTGAAAGGGC 8142  
 Qy 2622 LeuAlaCysSerIleGlySerProSerProAspProAlaLeuThrThrIleSerThr 2640  
 Db 8143 GTGCTCGCGCGAGAGAGCTGCACTGAGAGACTCCGCCACACAGGGCCACCTGCTG 8202  
 Qy 2641 ThrSerSerTyTrpAsnCysProSerProTyTrpAlaAspGly 2658  
 Db 8203 ACCGCGCTCCCTCAACTGCAACACCACTTCGTGAGCGGCTTGACATGCTGGCCACAGAC 8262  
 Qy 2659 TyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSer 2672  
 Db 8263 TTGGGGCAGATCCACCGCTCGCTGAGACATGTCAGGAGTGAAGGAGTCCAGAGAGCTC 8322  
 Qy 2673 2673 -----GlyIleSerGlnProSerTrpIlePro 2681  
 Db 8323 GGGGTGTCTCTGCGCTGAGAGGGGAGCGACAGAGAGAGCAGACCGCTCCCTCATGCCC 8382  
 Qy 2682 PheLeuLeuGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2701  
 Db 8383 AGAAGCTGCAAGAT 2701  
 Qy 2702 ProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAsp 2721  
 Db 8407 8407 -----CACGATTCGCACTCAGAT 8424  
 Qy 2722 SerAspLeuSerLeuGlnAspAspGlnSerGlySerTyTrpAlaSerThrHisSerSerAsp 2741  
 Db 8425 AGGAGGCTGTCCCTG---GATGAGCAGAGAGAGCTCTTACCCCTTCACACCTGTCAGAC 8481  
 Qy 2742 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2761  
 Db 8482 AGGAGAGAGAGATGGGGTGGAGGCTGAGGAGAAA-----TGG 8517

Qy 2762 AspSerLeuLeuGlyProGlyValAlaGlnArgLeuProLeuHisSerThrProIle 2780  
 Db 8518 GACCCGGCCAGAGGGC-----GCCGTCCACAGACACCCCAAGAGGAC 8559  
 Qy 2781 GlyIleProGlyProGlyValAlaProTrpProGlyAspPheGlyThrThrAlaValGln 2800  
 Db 8560 GCTGTGGCCACCAAGCTTCGGCGCGGTGGCCGACAGAGCTGCTGAGATGACAT 8619  
 Qy 2801 SerSerGlyAsnGlyValPro 2807  
 Db 8620 GAGAACCCCGCGCGCAAGCCCGCTGAAGTGGAGACCAAGTCAAGCTGAC 8679  
 Qy 2808 ---GlnGlnValLeuArgGlnAsnGlyAspAlaLeuSerArgGlySerLeuGlyPro 2826  
 Db 8680 CCGAGAGAGACAGAGGAGTCAACCTGAGAGTACCCCGGACACAGAGAGCGGGCGCA 8739  
 Qy 2827 LeuProGlySerSerAlaGlnPro-----HisIleGlyIleLeuIleValIleCys 2843  
 Db 8740 GCCAGGCTTGTAGAGAGCGAGCCCGACAGAGAGAGGATCTTGAATAATGAATC 8799  
 Qy 2844 Leu-----ProThrIleSerGlnIleSer-----SerLeuLeuArgLeuPro 2857  
 Db 8800 ACCTACCGCGCGCGCTGACGCTGACAGAGACAGCTGAGAGGCGCGCTCCGGAGAG 8859  
 Qy 2858 LeuGlnGlnCysThrGlySerSerArgIleSerSerAlaSerGlnGlySerArgIleGly 2877  
 Db 8860 CTGGCGCACTGTAGACAGAGCCCAATCTCGGCGCATCTCTCCGCGCTCTGGCGG 8919  
 Qy 2878 Pro-----ProProArgProProProArgGlnSerLeuGlnGlnGlnGln 2893  
 Db 8920 CCGAGCTGGCCATCAAGTCAAGAGCCCTGGAGAGAGCCCGGCGGTGACACCTTCAAC 8979  
 Qy 2894 GlyValMetProAlaMetSerIleValAlaGlnGlyThrValAspGlnAspSerSerGly 2913  
 Db 8980 GGGGTG-----GCCATGAATGTGCGCACTGGAGCGCCAGCGCATGCTCCAGC 9030  
 Qy 2914 SerGln 2915  
 Db 9031 TCTGAG 9036  
 RESULT 10  
 ABZ42827  
 ID ABZ42827 standard; DNA; 11389 BP.  
 XX  
 AC ABZ42827;  
 XX  
 DT 04-MAR-2003 (first entry)  
 XX  
 DE Human GPCR CELSR1/Flamingo nucleotide SRQ ID NO:443.  
 XX  
 KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor; modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200261087-A2.  
 XX  
 PD 08-AUG-2002.  
 XX  
 PF 19-DEC-2001; 2001WO-US50107.  
 XX  
 PR 19-DEC-2000; 2000US-257144P.  
 XX



Db 1312 CCGCTCAAGTCCACGGCCACCTGTGATCATCGAGTGGAGGACGAGAACCACTACCC 1371  
Qy 398 GlnPheSerGlnuValArgTyrValValGlnValArgGluuValThrProGlnAlaPro 417  
Db 1372 CAGTTCAAGCAGGAGCACTAGGTGTCAAGTCCCGAGGACGTGGGCTCAACACGGCT 1431  
Qy 418 ValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaValHisTyrSer 437  
Db 1432 GTGCTCGAGTGCAGGCGACGAGCCGGACCAAGGCGACAGACCGCGCATTCCTACAG 1491  
Qy 438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp 457  
Db 1492 ATCTCAGCGGGAACGTGGCCGCGCATGTTCTACTGCACTCGCTGACGGGAACTCCGAT 1551  
Qy 458 ValValSerProLeuAspTyrGlnThrThrThrGlyGlnTyrThrLeuArgValArgAlaGln 477  
Db 1552 GTGATCAACCCCTTGATTTGAGAGATGTCAGAAATACCTCGCTGAGCATTAAGGCCAG 1611  
Qy 478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
Db 1612 GATGGGGGCGGGCCCGCTCATCAATTTCTTCAAGGGGTGTCTGTGAGGTGGAT 1671  
Qy 498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer 517  
Db 1672 GTCAAGACACACAGCCTATCTTGTGAGCAGCCCTTCCAGGCGCACGGGTCTGAGAAAT 1731  
Qy 518 ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGluAspAsn 537  
Db 1732 GTGCCCTGGGCTACCCCGTGTGCACATTCAGGCGGTGACCGGACCTTGAGAGAAC 1791  
Qy 538 AlaArgLeuGlnTyrArgLeuAla-----GlyValGlyHis--- 549  
Db 1792 GCCCGGTGCACTATGCGCTGTGTGACACGGCTCCCACTTTCGGGGGCGGACGGCT 1851  
Qy 550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1852 GGGCTAAGAAATCCGCCCCACCCCTGACTTCCCTTCCAGATTCACACACACCTCGGT 1911  
Qy 561 TrpIleSerValAlaIleGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
Db 1912 TGGATCACTGTGTGTCCGAGCTGTGACCGGAGAGGTGAGACACTACGCTTCGGGGT 1971  
Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
Db 1972 GAGCGGTGACACCGCTGCGCCCGCCATGAGCTCCCTCACAGCGTGTCTACACGGT 2031  
Qy 601 LeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValAlaGluAsn 620  
Db 2032 CTGAGCTGATGACACACCGGTGTTCACGAGCCCACTACGAGCTTCTGTGAT 2091  
Qy 621 GluAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis 640  
Db 2092 GAGAGTCCGCGCGTGGGAGACAGGTGTCTGACCTCGAGGCCCGGCACTGACGCCAAC 2151  
Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
Db 2152 AGTGTGTTACTTACACGCTCACAGCGGCAACCCGGAACCGCTTGTGACCTCAGAGC 2211  
Qy 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGln 680  
Db 2212 CAGAGAGGGGCGGCTCATCATCCCTGCGTACTTGTGACCTGACGAGGAGGAGCGAG 2271  
Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
Db 2272 TAGCTGTGGCGGTGACGAGTCCGACGCGCACACGGTCCGACCTGCGATGCTTAATC 2331  
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheIleSerSerHisTyrThrValAsn 720  
Db 2332 AACCTCACTATGCAACCCACAGGCTGTCTTTCAGAGCTCCCACTTACACAGTGAAT 2391  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
Db 2392 GTCAAGTAGAGACGGCTGTGGGCACTTCATGCTACCTCACTGAGCCAAAGTAGAGAC 2451

Qy 741 ThrGlyIleAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACGAGAGGAATGCCCGCATCACTACGTGATTCAGAGCCCCGTGCCGATTCGCGATT 2511  
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
Db 2512 GACCCCGACAGTGGCACCATGATACAGATGAGAGCTGAGTATGAGAACCAAGGTCCGC 2571  
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db 2572 TACAGCTGACCATATGAGGCCCGAGCAACCGCATCCCGCAGAAATCAGACACCAACC 2631  
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db 2632 CTAGAGATCTCATATCTGATGAGCCAAATGACATGACCCCACTTCTGTGGATTTCTAC 2691  
Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGCTTCATCTTGAAGATGCTCACCTTCGACACACATCTCCAGGTCTTCCACAG 2751  
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyAspAspGly 860  
Db 2752 GACCGGAGCTCAGGTCCCAATGGGCTGTGTGACACTTCCAGGGTGGGAGCAGCGC 2811  
Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
Db 2812 GATGGGAGCTTCTATCATGAGCCACGTCCGGTGTGATTCGACCCACGCGCGGTGGAC 2871  
Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetPro--- 899  
Db 2872 CCGGAGAAATGTGGCGGTATCAACCTTGTGGGTCTGCGGTGTGAGTGGGCGGTCCACT 2931  
Qy 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
Db 2932 CCCCTTAGCGCTCGGTAAATTCAGAGTACATCTTGAGATTAATGACATGCCCC 2991  
Qy 920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTTTGAAGAGAGCACTGAGCTGTTGTGTGAGAGAACCAACCAAGTGGGTCCGG 3051  
Qy 940 ValAlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGCAAAATTCGTGTACACACCTGATGAAGCCCTTAATGCCAGATCAATGATACG 3111  
Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
Db 3112 ATTTGTGAAGGAGACATGCGGATTTCTTCAGCTGAGACTCTCAACGGGACCTGGCT 3171  
Qy 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
Db 3172 GCCATGTGAGACTGACTTGTGAGGTCCGGCGGAGTATGTCTGTGGTGCAGGCCAG 3231  
Qy 1000 SerAlaProLeuValSerThrAlaThrValHisValAlaGluLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGCTGTGATGAGCCAGGACAGGTGACACTTCTGTGTGACCAATATACAC 3291  
Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3292 CCGCTGTCTGCCGACTTCAGATCTCTTCAACAATATGTCAACCAAGTCCAAC 3351  
Qy 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGGGGTATCGGCTGATCCCGGCGCATGACCCGACCTGTCAACAGC 3411  
Qy 1060 LeuThrTyrSerPheGlnArgGlyAsnGluLeuSerIleValLeuLeuAsnAlaSerThr 1079  
Db 3412 CTCAACTACACTTGTGTGAGGGCAACAGCTGCGCTGTGTCTGTGTGACCCCGCACG 3471  
Qy 1080 GlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluValIleMetSer 1099  
Db 3472 GCGAATCTCACTAGCGCGACCTGTGAGCAACACCGCGCTGTGAGGCGCTCATGGAG 3531

1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 3532 GGTCTGTCTGTATGATGATCAACAGCTCAACGGCTTCTGCACTCTGTGTCACCATC 3591  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Db 3592 ATCAACGAGACAGCTGTGACCAACAGCACTCACTGCTGCGAGAACATGTCACGAG 3651  
Qy 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
Db 3652 AAGTTCCTGCTCCCGCTGCTGCGCTCTTCTGTGAGAGGGGTGGCCGCTGCTGTCAAC 3711  
Qy 1160 ProProAspHisValValValPheAsnValGlnHisAspThrAspAlaProGlyGlyHis 1179  
Db 3712 ACCAAGACGAGCGTCTTCTGCTTCAACGCTCAAGAACACACCGACCTC---AGCTTCAAC 3768  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPhe 1199  
Db 3769 ATCTGAACTGACCTTCTCGGCGCTGCTGCTGCGGCGCTCGCGC-----CAGTTC 3822  
Qy 1200 LeuProSerGluAspLeuGlnGlyArgLeuTyLeuAsnArgSerLeuLeuThrAlaIle 1219  
Db 3823 TTCCTGCGAGAGACTGCGAGGAGAGATCTACTGATCGACGCTGTCGACCAACATC 3882  
Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Db 3883 TTCAGCGACGCGGTGCTGCTTCAACGACCATCTGCTGCGGCGAGCTGCGAGAAC 3942  
Qy 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 3943 TACATGAATGCGGTGCTTCTTCCGATTCGACAGCTCCGCGCTTCTTCTCAAGCTCAC 4002  
Qy 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
Db 4003 ACCGAGCTCTTCCGCGCCATCCACCCCATCAACGCGCTGCGCTGCGCTGCGCGCGC 4062  
Qy 1280 PheThrGlyAspTyrCysGlnThrGlyValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 4063 TTCACCGGAGACTGACGAGAGAGATCGACTCTGTACTTCCGACCCGTCGCGCGC 4122  
Qy 1300 HisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 4123 AACGCGCGCTGCGCGAGCGCGAGCGCGCTACCTGCGAGTGTCTGAGAGACTTCACT 4182  
Qy 1330 GlyGluHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsn 1339  
Db 4183 GGAGAGACTGTGAGGTGATCCCGCTCAGGCGCGCTGTGCAACGAGGTGTGCAAGAAC 4242  
Qy 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAsp 1359  
Db 4243 GGGGCGACTGTGTAACCTGTCTCAACGCGCGCTTCCACTGCGTGTCTCTCGCGAG 4302  
Qy 1360 PheGluTyProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Db 4303 TATGAGAGCGCTTACTGTAGGTGACACACAGAGCTTCCGCCCACTTCTTCTGAC 4362  
Qy 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGlu 1399  
Db 4363 TTCGCGGCGCTGAGACAGCGCTTCCACTTCAACATCTCCCTCAGCTTGTGCAAGAA 4422  
Qy 1400 ArgAspArgLeuLeuLeuTyrAsnGlyArgPheAsnGlyHisIleAspPheValAlaLeu 1419  
Db 4423 AGGAACGCGCTTCTTCTTCAACGCGCGCTTCAATGAAACAGACATTCACCTCCG 4482  
Qy 1420 GluValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
Db 4483 GAGATGCTGAGAGAGAGGTGACGCTCACTTCTGCGAGCGAGAACAAACAGACCGTG 4542  
Qy 1440 SerProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuTyTyr 1459  
Db 4543 GCACCGAAGGTTCACAGTGTGTGATGACGCGGCGGTGCACTGTGTCAGAGTGCATAC 4602  
Qy 1460 TyrAsnLeuProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIleVal 1479

4603 TACAACAAGCCCAATATGCGCACCTGGCGCTGCCCAATGAGCGCTCGGAGAAAGATG 4662  
Qy 1480 AlaValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
Db 4663 GCGGTGTGACAGTGTGATGTGTGACAAACATGGCTGTGCGCTTGTGAAGAGATC 4722  
Qy 1500 GlyAsnTySerCysAlaAlaGlnGlyThrGlnGlyGlySerTyLeuSerLeuAspLeu 1519  
Db 4723 GGAATCTACAGTGTGCTGCCAGGAGCACTCAACCGGCTCAAGAAATCTCTGATGTG 4782  
Qy 1520 ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet 1539  
Db 4783 ACCGCGCTTCACTCTCGGCGGTGTCTCCCAACTGCGCAAGAAAGCTTCCAGTGCACAC 4842  
Qy 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
Db 4843 CGGAGTTCGTGGGTGTGATCGGAACCTGTCAAGTCAGCGCAAAATGTGACATGGCC 4902  
Qy 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaTyIleAsnValCysAsp 1579  
Db 4903 GGAATTCATCGCAACAAATGAGCACCGGAGAGCTGCGCTGCGAGGAACTTCTCGAT 4962  
Qy 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGlu 1599  
Db 4963 GGAAGCGGTGTCAAGATGAGAGCACTGTGTCAACAGTGTGAATGTATCTGTGTAG 5022  
Qy 1600 CysProLeuGlyPheGlyGlyLeuSerCysAlaGlnGlnMetAlaAsnProGlnHisPhe 1619  
Db 5023 TGTCCACTCCATTCGGGAGAAAGACTGTGAGAACGCAAGCTCCACCCAGCTTTC 5082  
Qy 1620 LeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyr 1639  
Db 5083 AGCGGTGAGAGCGTGTGCTGTGAGTGAAGCTGAACATCACTGTGTGCTGCTGCTG 5142  
Qy 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
Db 5143 CTGGGCTCATGTTCCGAGCCCGAGAGAGAGCAGCTTCAAGTGAAGAGCAGCAAGTGT 5202  
Qy 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGlnGly 1679  
Db 5203 GGGGCCACAGCTTCCCTTCAATCTGAATCTGAACAACTCACTCAAGTTGAGTGTCCAC 5262  
Qy 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTyr 1699  
Db 5263 GAGCCCTCCAGTGTGAGTCCGTGATGCTGCGGTTGCGGTTGACCGAGGAGATGG 5322  
Qy 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
Db 5323 CACCACTGCTGATCAGCTGAAGATGTTAAGAGAGACAGTGAATGAACACCTGTGTC 5382  
Qy 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
Db 5383 ACCATGACTTGGACTATGGATGTGACCAAGAACAGATATCGGGGGAGATCTTCCC 5442  
Qy 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyLeuProGlyProAlaGlyValAla 1755  
Db 5443 GGGCTGACGTAAGAGAGCGTGTGTCGAGGCGCTCTGAAGACAGATCTCGGTGGC 5502  
Qy 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsn 1775  
Db 5503 CGTGAATTCGAGGCTCATGACAGAGAGTGAAGATGGAGGGAGAGCCCAACAGTGGCC 5562  
Qy 1776 SerLeuAspProSerHisGlyGluSerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
Db 5563 ACCGTGAACATGAACAGACACTCAAGTCAAGGTGAAGAGACGCTGTGATGTGAGAGAC 5622  
Qy 1796 ProCysAspSerAspProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815  
Db 5623 CCTGTATCTCGAAGCCCTGTGCTCCCAATATAGCGCTGTGCAACGCGCTGTGAGAGATAC 5682  
Qy 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835

Db 5683 AGCTGGCTCTGTGACAAAGGCTACCTTGGAAATAACTGTGTGATGCTGTACCTGAAAC 5742  
Qy 1836 ProCysgluHhIsglnserValCysThrArgLysProSerAlaProHhIsgLYrThrCys 1855  
Db 5743 CCTGTGGAGAAATGAGGGGCGCTGTGGCGCTCCCGGCTCCCGGAGGGCTACGATGCG 5802  
Qy 1856 GluCyProProbaHhIryLeuGlyProLYrCysgluThrArgLLeaHhIProCysPro 1875  
Db 5803 GAACTGTGGCGCCAGTCACTAAGGACCGCTGTGTGAAGAAACAACCTTCACCTTCGTGCCCC 5862  
Qy 1876 ArgGLYrTrpTrpLysHhIProThrCysGLYProCysaHhIProValSerLysglYpHe 1895  
Db 5863 AGAGGCTGTGGGGGGAACCCGCTGTGTGACCTGTGCACCTGTCCCTGTGACAAAGCTTT 5922  
Qy 1896 AspProaHhIProbaHhIryLeuSerGlyLysCysHhIProValSerLysgluHhIryrArPro 1915  
Db 5923 GATTCGGACATGTATTAAGCAACAGGCGAGTGCACATGCAAGAGAAATTACTCAAGCTC 5982  
Qy 1916 ProGlySerProThrCysLeuLeuCysaHhIProThrGlySerLeuSerAlaVal 1935  
Db 5983 CTAGCCAGGACACTGTCTGTGCTGTGAGCTGTCTCCCATGAGCTTCCACAGCCGCACT 6042  
Qy 1936 CysAspProGluHhIProGlyLysCysProCysLysProGlyValLLeGlyArgGlnCysaHhI 1955  
Db 6043 TGGGACATGGCCACCGGGGAGTGTGCTGTGACCGCGGCTCATCGGCGCCAGTGCAC 6102  
Qy 1956 ArgCysaHhIProbaHhIryLeuValThrThrAsnGLYCysgluValaHhIryrAspSer 1975  
Db 6103 CGCTGGACAAACCGTTTGTGCGAGGTCAACAGCTCGCTGTGAAGTATCTCAAGTGGC 6162  
Qy 1976 CysProaHhIaLLeGlyHhIagLYrLeTrpProaHhIryrArgHhIryLeuProHhIa 1995  
Db 6163 TGTCCCAAGACATTTAGAGCGCGGATCTGTGTGACACAGACAAAGTTGGGGACGGGCT 6222  
Qy 1996 AlaAlaProCysProLysGLYSerPheGLYrThrAlaValArgHhIProCysaHhIryrAr 2015  
Db 6223 GCGGTGCAAGCTCTTAAGGATCCGTTGGAATGCGGTCCGACACTCGGCGGGGAGAG 6282  
Qy 2016 GLYrTrpLeuProbaHhIryLeuHhIProThrSerLLeThrPheSerGluLeuLysglY 2035  
Db 6283 GGGTGTGTGCGCCAGAGCTCTTAACTGTACCAACATCTCTTCGTCAGCTCAAGGGCC 6342  
Qy 2036 PheAlaGluArgLeuGlnArgaHhIryLeuHhIryLeuHhIryrArgSerGlnGlnLeu 2055  
Db 6343 ATGAATGAAGAGCTGACCCGCAATGAGAGCGAGTGAAGCGCGCCAGGGCTGTGACGCTG 6402  
Qy 2056 AlaLeuLeuLeuArgaHhIryrThrGlnHhIryrThrAlaValArgHhIryrPheGLYSerA 2075  
Db 6403 GTGAGGGCGCTGGGAGTGTACAGACACAGGGGACGCTTTGGCAATGACGTGCGC 6462  
Qy 2076 ValAlaTrpGlnLeuAlaThrArgLeuLeuAlaHhIryLeuSerThrGlnArgGlyPheGLY 2095  
Db 6463 ACCGCTTACAGAGCTGTGGGCGCAAGCTCTTACAGACACAGAGCTGTGAGCGGCTTCAAC 6522  
Qy 2096 LeuSerAlaThrGlnaHhIryrValHhIryrThrGlnaHhIryrValaGlySerAlaLeu 2115  
Db 6523 CTGGCAAGCAAGAGAGCGGCACTTCAAGAGAGCTCATCACTCGGCGAGCGGCTC 6582  
Qy 2116 LeuAspThrAlaAsnLysArgHhIryrTrpGluLeuLLeGlnGlnThrGlnGlyLysAla 2135  
Db 6583 CTGGCCCCAGCAACAGGGGGGCGTGGAGCAAGTCCAGCGAGGAGCGAGCGGCGCAAGCA 6642  
Qy 2136 TrpLeuLeuGlnHhIryrGluAlaTrpAlaSerAlaLeuAlaGlnaHhIryrThr 2155  
Db 6643 CAGTGTCTGGCGCTGTGAGGGCTACTTCAAGCAACGTGGCAAGCAAGTGGCGGAGAG 6702  
Qy 2156 TyrLeuSerProPheThrLLeValThrProaHhIryrLLeSerValaValaGlyLeuA 2175  
Db 6703 TACCTGGGCGCTTGTGATGTACAGCGGCAACATGATCTGTGTGTGTGACATCTTGAC 6762  
Qy 2176 LysGLYrAsnPheAlaGlyAlaLysLeuProArgTrpGluAlaLeuArgGlyGlnGlnPro 2195  
Db 6763 AAGTTCAACTTACGGGAGCGAGGCTCCCGGATTCGACACACATCATGAAAGATTCCCC 6822

Qy 2196 ProaHhIryrThrValLLeLeuProGlnserValPheArg-----Glu 2211  
Db 6823 AGGAGAGTGAAGTCTCTCGCTCTCTTCCAGCCGACCTTCTTCAAGCAACCTGAAGAAAA 6882  
Qy 2212 ThrProProValaValArgProHhIry-----ProGly 2222  
Db 6883 GAAAGCCCCCTGTGAGGCGAGCTGTGGCGAGAGAACCCCGGACCAAGCGGCGCGGAG 6942  
Qy 2223 GluAlaGlnGluProGluGluLeuAlaArgArgGlnArgaHhIryrProGluLeuSerGln 2242  
Db 6943 CTGGCAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7002  
Qy 2243 GlyLysAlaValaLLeSerValLLeLLeTrpArgThrLeuAlaGlyLeuLeuProHhIry 2262  
Db 7003 CAGTTTGGCGGTGCTGTGTGATCATTTAACGCAACCTTGGGAGAGCTCTGTGCGGAGGCG 7062  
Qy 2263 TyrAspProaHhIryrSerLeuArgValProLysArgProLLeLLeaHhIryrProVal 2282  
Db 7063 TACGACCCCGAGCCGTGAGCGCTCGGTGTGCTCAACCGCCCATATTAATACCCGAGAG 7122  
Qy 2283 ValSerLLeSerValHhIryAspArgGluLeuLeuProArgAlaLeuAspLysProVal 2302  
Db 7123 GTGACACAGCTGTGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7182  
Qy 2303 ThrValGlnPheArgLeuLeuGlnThrGluArgTrpHhIryrProLLeCysValaPheTrp 2322  
Db 7183 CTGTGTGAGTGTGCGCTGTGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7242  
Qy 2322 AsnHhIryrLLeLeuValSerGlyrThrGlyrTrpSerAlaArgGLYCysgluValaVal 2342  
Db 7243 AACCACTCCCTGGCCGTGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7302  
Qy 2343 PheArgaHhIryrSerValaSerCysGlnCysaHhIryrMetThrSerPheAlaValLeu 2362  
Db 7303 TCCAGGAACCGGACATGTGCGCTGTGACAGTGCACCAACAGCGAGCTTGTGGGTGCTC 7362  
Qy 2363 MetAspValSerArgArgGluHhIryrLLeLeuProLeuLysTrpThrLeuThrTrpVal 2382  
Db 7363 ATGGAATCTCCAGCGCGAGAAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7422  
Qy 2383 AlaLeuGlyValaLLeLeuAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArg 2402  
Db 7423 GCTGTGTCTTCACTGTGAGAGCTGTGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7482  
Qy 2403 LLeLeuArgSerAsnGlnHhIryrLLeArgArgaHhIryrAlaLeuGlyLeuAla 2422  
Db 7483 AGCTGTGCTTCACTGTGAGAGCTGTGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7542  
Qy 2423 GlnLeuValaPheLeuLeuGlyrLLeaHhIryrAlaAspLeuProPheAlaCysThrValLLe 2442  
Db 7543 CAGCTGT 7602  
Qy 2443 AlaLLeLeuLeuHhIryrPheLeuTrpCysThrPheSerTrpAlaLeuLeuGlnLLeu 2462  
Db 7603 GCGATCTCTTCACTGT 7662  
Qy 2463 HhIryrTrpArgAlaLeuThrGluValaArgAspValaLeuThrGlyrProMetArgPheTr 2482  
Db 7663 CATGTCTACCGGATGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7722  
Qy 2483 TyrMetLeuGlyrTrpGlyValaProAlaPheLLeThrGlyLeuAlaValaGlyLeuAspPro 2502  
Db 7723 TACGT 7782  
Qy 2503 GluGlyrTrpGlyrAspProaHhIryrCysTrpLeuSerLLeLLeTrpThrLeuLLeTrpSer 2522  
Db 7783 CAGGCGTACGAGAACCCGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7842  
Qy 2523 PheAlaGlyrProValaLLePheAlaValaSerMetSerValaPheLeuTrpLLeLeuAla 2542  
Db 7843 TTTGGGGGGCCCATGAGAGCTGTATTAATCATCAACAGTCACTTGTGTCTTATCTGCA 7902









QY 1179 Hs11eLuanValSerLeuSerValGlyInProProGlyProGlyGlyProPro 1188  
 Db 3769 AACATCTGGAACGTGACCTTCTCGCGCTGCTGCGCGGCTCGCGGC-----CAG 3822  
 QY PheLeuProSerGluAAspLeuGlnGluArgLeuYrLeuAAspArgSerLeuThrAla 1218  
 Db 1199 TTTCTTCCCGGAGAGACCTGACGAGAGACGATCTACGAAATCGAACCTGCTGACACC 3882  
 QY 1219 ILeSerAlaGlnArgValLeuProPheAAspAAsn1LeCysLeuArgGluProCysGlu 1238  
 Db 3883 ATCTCAGACGACCGGCTGCTGCTTCCAGACGAAATCTGCTGCGGAGCCCTGCGCAG 3942  
 QY 1229 AAsnYrMetArgCysValSerValLeuArgPheAAspSerAlaProPhe1LeAlaSer 1258  
 Db 3943 AACTACAGAAAGTGGTGTCTGCTGCGATTCGACACTCGCGCCCTTCTCAGCTCC 4002  
 QY 1259 SerSerValLeuPheArgPro1LeH1sProValGlyLeuArgCysArgCysProPro 1278  
 Db 4003 ACCACCGTCTTCTTCCGCGCCATCCACCCCAACAGGCTGCGCTGCGCGCTCGCGCC 4062  
 QY 1279 GlyPheThrGlyAspYrCysGluThrGluValAAspLeuCysYrSerArgProCysGly 1298  
 Db 4063 GGCCTTCCACCGGAGACTTACGAGAGACGAGATCGACTCTGCTACTCCGACCCGCGCGC 4122  
 QY 1299 ProH1sGlyArgCysArgSerArgGluGlyYrYrThrCysLeuCysArgAspGlyYr 1318  
 Db 4123 GCCAACCGGCGCTGCGGACCGCGAGCGGCGCTACCTGCGAGTGTCTTCGAGGACTTC 4182  
 QY 1319 ThrGlyGluH1sCysGluValSerAlaArgSerGlyArgCysYrThrProGlyValCysLeu 1338  
 Db 4183 ACTGAGAGACCTGTGAGAGTGAATGCCGCTCAGCGCTGCGCCATCCAGCGGAGTGGAG 4242  
 QY 1339 AAsnGlyYrThrCysValAAsnLeuValGlyGlyPheYrCysAspCysProSerGly 1358  
 Db 4243 AACGGGGGACCTGCGTGAACCTGCTCAACGCGGCTTCCACTGCGTGTCTCTCGGC 4302  
 QY 1359 AspPheGluYrProYrCysGlnValThrThrArgSerPheProAlaH1sSerPhe1Le 1378  
 Db 4303 GAGATGAGAGGCGCTTACTGTGAGGTGACACAGAGACTTCCGCGCCAGCTCTCTGTC 4362  
 QY 1379 ThrPheArgGlyLeuArgGlnArgPheH1sPheThrLeuAlaLeuSerPheAlaThrYr 1398  
 Db 4363 ACTTCCGGGCGCTGAGACAGCGCTTCCACTTCAACATCTCTCCATCGTTGCCACTCAG 4422  
 QY 1399 GluArgAspGlyLeuLeuLeuYrAAsnGlyArgPheAAsnGluYrH1sAspPheValAla 1418  
 Db 4423 GAAAGAAACGGCTGCTCTCTCAACAGCGGCTTCAATGAGAGACAGACTCATCGCC 4482  
 QY 1419 LeuGluVal1LeGlnGlnGlnValGlnLeuThrPheSer-----AlaGlyGluSerThr 1436  
 Db 4483 CTGAGATCTGTGACGAGACAGGTGTGACTCTCTCTGAGGTGAGCGGAGACAAACA 4542  
 QY 1437 ThrThrValSerProPheValProGlyValSerAspGlyGlnThrPhe1sThrValGln 1456  
 Db 4543 ACGACCGTGGGACCGAAGTTCCTCAGGTGTGTGATGACGGGCGGTGCACTCTGTGAG 4602  
 QY 1457 LeuYrYrYrAAsnYr-----ProLeu 1464  
 Db 4603 GTGCGACTACTCAACAAGTAAGATGGGCCCAACCACTTCCCTGCGCCCAACCCCAAT 4662  
 QY 1465 LeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnYrValAlaValAlaThrVal 1484  
 Db 4663 ATTGCCACCTGGGCGCTGCCATGGGCGCTGCGGAGAAAGATGGCGGTGTGACAGTG 4722  
 QY 1485 AspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGlyAAsnYrSerCys 1504  
 Db 4723 GATGATGTGACAAACATGGCTGTGTGCTTTGGAAAGACATCGGGAATACAGCTGC 4782  
 QY 1505 AlaAlaGlnGlyThrGlnGlyGlySerYrYrYrSerLeuAAspLeuThrGlyProLeuLeu 1524  
 Db 4783 GCTGCCCGGAGCACTCAACCGGCTCCAAAGAGTCCCTGATTCGACCGGCGCTCTACTC 4842  
 QY 1525 LeuGlyGlyValProAspLeuProGluSerPheProValArgMetArgGlnPheValGly 1544

Db 4843 CTGGGGGAGTGTCCCAACCTGCGCAAGAAAGCTCCCAAGTGCACAAACGGGAGTGTGAGGC 4902  
 QY 1545 CysMetArgAAsnLeuGlnValAAspSerArgH1sAlaSerMetAlaAspPhe1LeAlaAsn 1564  
 Db 4903 TGCATGGGAACCTGTGACGTGACCGGCAAAATGGACATGGCGGAGTTCATCGCCAAC 4962  
 QY 1565 AAsnGlyThrValProGlyCysProAlaYrYrAAsnValCysAAspSerAAsnThrCysH1s 1584  
 Db 4963 AATGGACCCGGGAAGGCTCGCGCTCGAGAGAACTTCTGCAATGGAGAGCGGTGTGAG 5022  
 QY 1585 AAsnGlyYrThrCysValAAsnGlnThrAAspAlaPheSerCysGluCysProLeuGlyPhe 1604  
 Db 5023 AATGAGGACCTGTGTCAACAGGTGAATATATCTGTGTGATGATCTTCACCTCCGATTC 5082  
 QY 1605 GlyGlyYrSerCysValAAsnGlnLeuMetAlaAsnProGlnH1sPheLeuGlySerSerLeu 1624  
 Db 5083 GCGGGAGAACTGTGAGCAAGCCATGCTTCAACCCCAAGCTTTCAGCGGTGAGAGGCTC 5142  
 QY 1625 ValAlaThrH1sGlyLeuSerLeuPro1LeSerGlnProTrpYrLeuSerLeuMetPhe 1644  
 Db 5143 GTGCTCGAGTGAACCTGAACATCATCATCTGTGTGCTGCTGCTGCTGCTGCTGCTGCT 5202  
 QY 1645 ArgThrArgGlnAlaAspGlyValLeuLeuGlnAla1LeThrArgGlyArgSerThr1Le 1664  
 Db 5203 CGGACCCGGAAGAGACAGCGTTCGTGATGAGAGCCACAGGTGTGGGCCCAAGCTTT 5262  
 QY 1665 ThrLeuGlnLeuArgGluGluYrH1sValMetLeuSerValGluGlyThrGlyLeuGlnAla 1684  
 Db 5263 CGCTTCAATCTCTGAACACTACTTCCAGTTTGAAGTGTCCACGCGCCCTCCGATGTG 5322  
 QY 1685 SerSerLeuArgLeuGluProGlyArgAlaAAsnAspGlyAspTrpH1sH1sAlaGlnLeu 1704  
 Db 5323 GAGTCCGTGAGTGTGTGCTCGGAGTGGAGGAGACCGAGAGAGTGGACCACTGCTGATC 5382  
 QY 1705 AlaLeu-----GlyAlaSerGlyGlyProGlyYrH1sAla1Le-----LeuSerPheAsp 1720  
 Db 5383 GACGTGAAGATGTTAAGAGAGACAGTGAATGAGACACTGTGTCAACATGACCTTGGAC 5442  
 QY 1721 TyrGlyGlnGlnArgAlaGluGluAAsnLeuGlyProArgLeuH1sGlyLeuH1sLeuSer 1740  
 Db 5443 TATGTGATGACCAAGAAACAGAGCATATCGGGGAGATGTTCCCGGCTGACAGGTAAAG 5502  
 QY 1741 AAsn1LeThrValGlyGly1LeProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
 Db 5503 AGCGTGTGTGTGAGAGCGCGCTCTGAAGACAGGTCTCGGTGCGCGTGTATCCAGGC 5562  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAAsnSerLeuAspProSer 1780  
 Db 5563 TGCATGAGAGAGTGAAGATGGGGGAGCGCCACCAACGTCCACCTGAACATGAAAC 5622  
 QY 1781 H1sGlyGluSer1LeAAsnValGluGlnGlyCysSerLeuProAspProCysAAspSerAAsn 1800  
 Db 5623 AAGGACATCAAGGTACAGGTGAAGAGACAGCTGTGATGTGACAAACCTGTACTGAGC 5682  
 QY 1801 ProCysProAlaAAsnSerYrCysSerAAsnAspTrpAspSerYrYrYrSerCysYrAsp 1820  
 Db 5683 CCTGTGCTCCCAATAGCGCTGCGACAGCGCTGAGAGACTACAGCTGCGTGTGAC 5742  
 QY 1821 ProGlyYrYrYrYrYrAspAsnYrThrAAsnValCysAAspLeuAsnProCysGlnH1sGln 1840  
 Db 5743 AAAGGATCTTGAATTAACGTGTGATGCTGTGACCTTGAACCTTGCAGAAACATG 5802  
 QY 1841 SerValCysThrArgYrAspSerAlaProH1sGlyYrThrCysGluCysProProAla 1860  
 Db 5803 GGGGCGTGGGTGCTCCCGCGCTCCCGGAGGCTTACGTGTGAGAGTGTGGGCCAAT 5862  
 QY 1861 TyrLeuGlyProYrCysGluThrArgGlyAAspGlnProCysProArgGlyYrTrpGly 1880  
 Db 5863 CACTTACGGGCGGTATCTGTAGAAACAACTGCATTCCTGCGCCGAGAGCGGTGTGGGG 5922  
 QY 1881 H1sProThrCysGlyProCysAAsnCysAAspValSerYrYrPheAspProAspCysAAsn 1900



```

QY 2608 SerTyrValValLeuSerTyrGluValArgLysAlaLeuLys---LeuAlaCysSerArg 2626
DB 8143 TTCACCTGCTGCTCAACGACGAGAGTCCGGAAGCACTGAAGGGCGTGCCTCGGGGAGG 8202
QY 2627 LysProSerProAlaLeuThrThrLysSerThr---LeuThrSerSerTyrAsn 2645
DB 8203 AAGCTGACCTGGAGAGTCCGACCAACAGGGCCACTGTCAGAGCCCTCCCTCCAC 8262
QY 2646 CysProSerProTyrAlaAspGly-----ArgLeuTyrGlnProTyrGlyAspSerAla 2663
DB 8263 TGCACACACACCTTCGCTGACGGGCTGACATGCTGGACACAGACTGGGCGAGTCCACC 8322
QY 2664 GlySerLeuHisSerThrSerArgSer----- 2672
DB 8323 GCCTCGCTGGACATGTCAGAGGATGAAGGATCCAGAGCTGGCTGCTCTGGG 8382
QY 2673 -----GlyLysSerGlnProSerTyrTleProPheLeuLeuArgGlu 2686
DB 8383 CTGGTGAAGGGGACCCAGAGAGCCAGACCGCTCCCTCATGCTCCAGAGCTGCAAGAT 8442
QY 2687 GluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeuPhe 2706
DB 8443 -----CCCCCTGGC----- 8451
QY 2707 LeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeu 2726
DB 8452 -----CACGATTCACGATCCAGATCAGATCAGAGCTGTCCCTG 8484
QY 2727 GluAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlu 2746
DB 8485 ---GATGAGGACAGACCTTACGCTTACGCTCCTCAGACCTGTCAGACGAGAGAGATGGG 8541
QY 2747 GlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGlnGlnGlnGlnGlnGly 2766
DB 8542 GTGGAGACTGAGGAAAA-----TGGAGACCGGGCCAGGGGCG 8577
QY 2767 ProGlyAlaGluArgLeuProLeuHisSerThrProLys---AspGlyGlyProGlyPro 2785
DB 8578 -----GCCGTCCACAGCACCCCAAGGGGCGCTGTGGCCACAC 8619
QY 2786 GlyLysAlaProTyrProGlyAspPheGlyThrThrAlaLysGlnSerSerGlyAsnGly 2805
DB 8620 GTTCCGGCCGCTGAGCCGACCAAGAGCTGCTGAGATGACAGTACAGACCCCAAGGGCG 8679
QY 2806 AlaPro-----GlnGluArgLeu 2811
DB 8680 AAGCCCGGCTGAAGGTGAGACCAAGGTGAGCTGAGCTGACCGCGAGGAGCAGGGC 8739
QY 2812 ArgGlnAsnGlyAspAlaLeuSerArgGlnGlySerLeuGlyProLeuProGlySerSer 2831
DB 8740 AGTCACCGTGAAGTACCCCGGACCAAGAGCGGGGCGGACGCAAGCTTGCTAGC 8799
QY 2832 AlaGlnPro-----HisLysGlyLysLeuLysLysCysLeu-----Pro 2845
DB 8800 AGCCAGCCCCCAGAGGACGAGAGGACATCTTGAATAATTAAGTACACTTCCGCGCGCTG 8859
QY 2846 ThrLysSerGlyLysSer-----SerLeuLeuArgLeuProLeuGlnGlnGlnGlnGly 2863
DB 8860 ACGCTGACGAGGACGAGCTGAGAGGCGCGCTCCGGGAGAACTGGCGCACTGTAGACAG 8919
QY 2864 SerSerArgGlySerSerAlaSerGlnGlySerArgGlyPro-----Pro 2879
DB 8920 AGCCCAACATCCTCGGACGCTCTTCCCTGCGCTGCGCGGCGCCAGCTGCGCCATCACA 8979
QY 2880 ProArgProProProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2899
DB 8980 GTCAAGAGCCCTGGAGAGGAGACCGGGGCGTACCACTCAACGGGGTG-----GCC 9030
QY 2900 MetSerLysAlaGlyThrValAspGluAspSerSerGlySerGln 2915
DB 9031 ATGATATGCGCACTGGAGAGCGCCAGCGCATGTGCTCCGACTCTAG 9078

```

```

AAT85320
ID AAT85320 standard; cDNA; 10195 BP.
XX
AC AAT85320;
XX
XX 08-DEC-1997 (first entry)
XX
DE Mouse receptor ME2 encoding cDNA.
XX
XX Chemical screening; epidermal growth factor; EGF; repeat; cadherin;
XX neuronal cell division; differentiation; cell physiology; trauma;
XX neurodegenerative disease; adenylate cyclase; cAMP; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 2..8125
FT /tag=a
FT /product= ME2
XX
XX MO9707209-A2.
XX
XX 27-FEB-1997.
XX
XX 14-AUG-1996; 96WO-GB01980.
XX
XX 21-AUG-1995; 95GB-0017098.
XX
XX (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
XX Hadjantonakis A, Little P,
XX
XX WPI; 1997-165293/15.
XX
XX P-PSDB; AAW27161.
XX
XX Screening chemicals which interact with receptor contg. EGF-like
XX and/or cadherin-like repeats - to develop pharmaceuticals for
XX treating neurodegeneration, trauma, etc.
XX
XX Claim 12; Fig 9; 55pp; English.
XX
XX A method has been developed for the screening of a chemical for
XX subsequent use as a pharmaceutical agent. The method involves contacting
XX the chemical with a receptor comprising epidermal growth factor
XX (EGF)-like repeats and/or cadherin-like repeats and determining whether
XX the chemical interacts with the receptor to form a complex. The present
XX method encodes the mouse receptor ME2, which can be used in the above
XX method. The method is used to screen for chemicals which can cause
XX neural cells to divide or differentiate, can affect cellular physiology,
XX repair trauma, treat neurodegenerative disease, stimulate adenylate
XX cyclase production, increase cAMP levels and promote neural growth.
XX
XX Sequence 10195 BP; 2265 A; 2981 C; 2855 G; 2094 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 10195
XX Score: 8710.00 Matches: 1668
XX Percent Similarity: 74.41% Conservative: 408
XX Best Local Similarity: 59.78% Mismatches: 601
XX Query Match: 56.03% Indels: 116
XX DB: 18 Gaps: 29
XX
XX US-09-916-849A-3 (1-2923) x AAT85320 (1-10195)
XX
XX 192 GlnAsnGlnProAlaGlyThrProValAlaSerLeuArgAlaLeuAspProAspGlnGly 211
XX 104 GAACGGAACCGTGGGAGCTGCGGTCACTGCAATGCGGCGGACACCAAGAGG 163
QY 212 GlnAlaGlyArgLeuGlnGlyThrThrAspAlaLeuPheAspSerArgSerGlnGlnPhe 231
DB 164 GATGACAGAGCGCTTGTGATACCAAGATGAGAGCGGCTGTGATGAGCGCTTAATGCTAC 223
QY 232 PheSerLeuAspProValThrGlyAlaValThrThrAlaGlnGlnGlnGlnGlnGlnGln 251

```

```

Db      224  TTCTCATGATGATCGCCGACGAGGTGACGACACGCCGCTCCCTGACCGGAGAAC 283
Qy      252  LysSerThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAla 271
Db      284  AAGACACGTCAGTACTCAAGTTAGTGTGTGGACCAAGCTCCCGAGGGGCTCAGCT 343
Qy      272  LeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGln 291
Db      344  GCCACCTACCTCACTGACAGTGAACATAACCAACGACCCAGCTCTTGAGCAG 403
Qy      292  GlnGluTyrLeuGlnSerLeuArgGluAsnLeuGluValGlyTyrGlnValLeuThrVal 311
Db      404  TCTGAGTATCGAGAGGAAATTCGAGAAACCTGAGAGTGGCTATAGAGGTCTGACATC 463
Qy      312  ArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuGluGly 331
Db      464  CGTGGCACCAGCGGGGATGCCCTTCCAAACGCAACATGCGCTATGCTGTCTGGAGGGC 523
Qy      332  SerGlyGlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThr 351
Db      524  GCAAGGTGGT-----GCTTTGAGATAGACGACGATCAGAGTGTGTCGCGCAC 571
Qy      352  ArgGlyProValAspArgGluGlnValGluSerTyrGlnIleuThrValGlnAlaSerAsp 371
Db      572  CGAGCTGTGTGACCGGTAGAGAGCGGCTGAGTACCAAGCTCTGTGTGAGGCGCAATAG 631
Qy      372  GlnGlyArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAsp 391
Db      632  CAGGCTCGGAATCCAGGGCCCACTCAGTGTCTCAGCCCTCACAATAGTGTAAAGAC 691
Qy      392  AspAsnAspAsnAlaProGlnPheSerGluValArgTyrValIleGlnValArgGluAsp 411
Db      692  GAGAAATGACAACTACCCCACTTCACTGAGAAAGCCCTATGTGTCTCAAGTCCCAAGAAC 751
Qy      412  ValThrProGlyAlaProValLeuArgValThrAlaSerAspAspAspGlySerAsn 431
Db      752  GTAGCCGTCAACACGAGCTGTGTGAGTCCAGGSCCACTGACCGGAGACCAAGGGAGAT 811
Qy      432  AlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAla 451
Db      812  GCAGCCATACACTACAGCATGTGTAGTGCACCTGAAGGGGTCAAGTTCACTGCATTCG 871
Qy      452  GlnThrGlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGluTyrThr 471
Db      872  CTTAGTGGAGCTGATGTATCAACCGCTGACCTTCGAAGCCATCCGGGAAATACACC 931
Qy      472  LeuArgValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuVal 491
Db      932  CTGGCGATCAAAAGCCCAAGATGGGGCGGGCTCTCTCATTAATTCCTCAGAGACTGTG 991
Qy      492  ThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGln 511
Db      992  TCGGTGACAGGTGTGATGTGAACGACATGGCCCATCTTTGTAGAGAGCCCTTTCAG 1051
Qy      512  AlaThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAsp 531
Db      1052  GGTGCGGTGCTAGAAATGTGCCCTCGGCCCACTGACATCTCTGCACATCCAAAGCGGTGAC 1111
Qy      532  AlaAspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHis----- 549
Db      1112  GCAGATGACAGGGAGAAACCCAGGCTGACGTACCGTCTAGTGACACAGCTCCCATATC 1171
Qy      550  -----AspPheProPheThr 554
Db      1172  GTGGGGGGGACAGTGTGACTCTGAGAACCTGTGCTGCGCCAGACTTCCCTTCCAA 1231
Qy      555  IleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluValAsp 574
Db      1232  ATCCCAACAACGCTCGGTTGATCTGTGTGCGGAGCTGGAACCGTGAAGAGGTGAA 1291
Qy      575  PheTyrSerPheGlyValGluIleArgAspHisGlyThrProAlaLeuThrAlaSerAla 594

```

---

```

Db      1292  CACTATACCTTTGAGATGAAGACAGTGGACCAATGGCTTCAACCAAGCATAGCTCTTGCC 1351
Qy      595  SerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnProGlu 614
Db      1352  AAGGTCTCATCAAGCTCTGATGTATAATGATTAACAGACCCATTTCAACGAGCTGTG 1411
Qy      615  TyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSerAla 634
Db      1412  TATGAGCTGCGCTGAAATGAGATGCGGCTGTGCGGAGACAGCTGTACCTCAGAGGCC 1471
Qy      635  ValAspAspAspAlaHisSerValIlePheTyrGlnIleThrSerGlyAsnThrArgAsn 654
Db      1472  CGAGACCGTATGCGCAATATGTATATCACTTACCAAGCTGACGCGGTGGAAACACCGGCAC 1531
Qy      655  ArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAsp 674
Db      1532  CGCTTGGACATCAGACAGCCAGACGCGCGGTGGCTTATCACTTGACCTGCGCTTGAC 1591
Qy      675  TyrIleLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyTyrArgGluAsp 694
Db      1592  TACAAGCAGGAACGCGATATGTGTGCTGTGACCGGCTCGATGGCAGCGCTTCAAC 1651
Qy      695  ThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisArgProValPheGlnSer 714
Db      1652  ACCGCGAGGTCTTTATCAAGATTACAGATGCCAAACCCACAGGCGGTTTCCAGAGT 1711
Qy      715  SerHisTyrThrValAsnValAsnValAsnAspArgProAlaGlyThrThrValValLeuIle 734
Db      1712  TCCCACTTCAAGCTCAAGTGTAGTGAAGACCGGCGCGGTGGGCACTTCATGCTACATC 1771
Qy      735  SerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSer 754
Db      1772  AGTGCCAGATGAGATACCGGTGAGAACCGCGCATCACTATGTGCTAGAGATCCC 1831
Qy      755  IleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAsp 774
Db      1832  GTACCAAGTTCGCGATTTGACCCGACCTGGACACCTTATTAACATGACGAAATGGAC 1891
Qy      775  TyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGln 794
Db      1892  TATGAGACCAAGCGTGCTTACAGCTGGCCATCAAGCTCGAGCAATAGGATTCCTCAG 1951
Qy      795  LysSerAspThrThrTyrLeuGluIleLeuValLeuAspValAsnAspAsnAlaProGln 814
Db      1952  AAGTCAGACATCACTCTCTGAGATCTTTATCTCGACCGCAATGACAAAGCGCCGAG 2011
Qy      815  PheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerVal 834
Db      2012  TTCTGCGAGATTCTTACAGAGGTTCTGTTTTCAGAGATGCCCCCACTTACAGATGTC 2071
Qy      835  LeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPhe 854
Db      2072  CTCACAGTCTGTGTACAGACCGTACAGCGCCCTATATGGCGCGCTCTGTACACTTTC 2131
Qy      855  GlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleValArg 874
Db      2132  CAGGTGGGATGATGAGATGAGATTTCTACATTTGAGCCCAAGTGTGTGATTCGCT 2191
Qy      875  ThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuAlaGlyAlaTyrAlaVal 894
Db      2192  ACCAGCGCCCGGCTGACAGAGAAATGTGGCCGTGTACCAACTTTGGGCTCTGCGTGTG 2251
Qy      895  AspIleGlyMetPro---ProAlaArgThrProMetGluValThrValThrValLeuAsp 913
Db      2252  GATCGGGGAGCCCGAATCCCTCAGTGCCTCAGTGGAAATTCAGGTGAGGTGTGGAC 2311
Qy      914  ValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGlnGluAsn 933
Db      2312  ATTAACACAACACCCCAAGTGTGAGAAAGCAAGCTGAGCGTGTGTGGAAGAGAAC 2371
Qy      934  SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn 953
Db      2372  AGCCCTGTGGGTTCAGTGTGACAAAGATTAAGGCCCAACAGACCGGAGCAAGATCCGAT 2431

```



QY 954 AAGGlnIleMetYrGlnIleValGluGlyAsnIleProGluValPheGlnIleuAspIle 973  
 Db 2432 GCTCAAGTATTTATCATGATCGTGAAGGCAATGTGCCAGAGGCTTCCAGCTGGACCTA 2491  
 QY 974 PheSerGlyLeuLeuThrAlaLeuValAspLeuAspYrGluAspArgProGluYrVal 993  
 Db 2492 CTGAGTGTGATCCCTCGTCCCTGTGTGAGTGTGATTTGAGGTCGAGAGGACTATATG 2551  
 QY 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1013  
 Db 2552 TTGGTGTGAGGAGGCACTGTCTCTCTGTAGCCGGGCAACCGTGCACATCCGCTCTC 2611  
 QY 1014 LeuAspArgAsnAspAsnProProValIleuGlyAsnPheGluIleLeuPheAsnAspYr 1033  
 Db 2612 CTGGACCAAGATGACCAACCAACCGAGTGTGCTCACTTCCAGATCTTTCAACAATAT 2671  
 QY 1034 ValThrAsnArgSerSerSerPheProGlyAlaAlaIleGlyArgValProAlaHisAsp 1053  
 Db 2672 GTCAACCAATTAATTCACAGCTTCCCAAGTGTGTGATGAGCGGCATCCAGCCCAAGAC 2731  
 QY 1054 ProAspIleSerAspSerIleuThrYrSerPheGluArgGlyAsnGluLeuSerLeuVal 1073  
 Db 2732 CTGACCTATCTGACACCTCTAATTACCTTTCTGCAAGGCAACGAGCTGAGCTGTCTG 2791  
 QY 1074 LeuLeuAsnAlaSerThrGlyGlyLeuLeuLeuSerArgAlaLeuAspAsnAsnArgPro 1093  
 Db 2792 CTGTGTGATCCCGCCAGAGAGAGTTCAGCTCAGCGGAGTCTGCAACAACCCGCGCA 2851  
 QY 1094 LeuGlnAlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGlnCys 1113  
 Db 2852 CTGAGGCGCTCAGAGAGGTGTGTGTGTGATGTGATCCACAGCCGCTCTCTGTGC 2911  
 QY 1114 AlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeu 1133  
 Db 2912 ACTGTGGCGGTGACCATTTACAGATGACATGTGTGACCAAGCAATCATGTCCGCTG 2971  
 QY 1134 GluAspMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaVal 1153  
 Db 2972 GAGAAACATGTGCGAGAGAGTCTGTCCCGCTGTCTCCCTTTGAGAAAGGGGTG 3031  
 QY 1154 AlaAlaThrLeuAlaThrProProAspHisValAlaValPheAsnValGlnArgAspThr 1173  
 Db 3032 GCCACAGTACTGTCCACCAACAGAGATGACATCTTCGTTCAACATCCAGAAACAGACG 3091  
 QY 1174 AspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyPro 1193  
 Db 3092 GAGGTCTC---AGCTCCACATCTGACATGATCTTCTCGGACATGCTCCCGGTGGACCC 3148  
 QY 1194 GlyGlyGlyProProPheLeuProSerGluAspLeuGlnGluArgLeuYrLeuAsnArg 1213  
 Db 3149 CGTGGC-----CGGTCTTCTCCCGTCAGAGGACCTGAGAGGACATCTTAACCGG 3202  
 QY 1214 SerLeuLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsnAlaIleCysLeu 1233  
 Db 3203 ACATGTGCTCACCACTTCCTCGCCAGAGGTGTGTGCTTGTATGACAAATGTGCTG 3262  
 QY 1234 ArgGluProCysGluAsnYrMetArgCysValSerValLeuArgPheAspSerSerAla 1253  
 Db 3263 AGGAGAGCCCTGCGAAGTACATGAGTGTGCTGCGGTGCTTGTGACAGTTTGGCA 3322  
 QY 1254 ProPheIleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArg 1273  
 Db 3323 CCTTCATTTAGTTTCACACGAGTGTCTTTCGGGCTTCCATCCATCAACGGGCTTCGCG 3382  
 QY 1274 CysArgCysProProGlyPheThrGlyAspYrCysGluThrGluValAspLeuCysYr 1293  
 Db 3383 TGCCGCTGCGCGCGGGTTTACCGGGAGCTACTGCGAGCTGAATTAATGATCTTGTCTAC 3442  
 QY 1294 SerArgProCysGlyProHisGlyArgCysArgSerArgGluGlyYrYrThrCysLeu 1313  
 Db 3443 TCCATATCGTGGGCGGCAATGCGGCTGTGCGAGAGCCGTGAGGTGGCTACACTTGTGAG 3502

QY 1314 CysArgAspGlyYrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysYr 1333  
 Db 3503 TGCTTGAAGACCTTCACTGTGGAGCATTTGCCAGTGAACGTTTCGCTCAGGCGGCTGTGCC 3562  
 QY 1334 ProGlyValCysLeuAsnGlyGlyYrThrCysValAsnLeuLeuValGlyGlyPheCys 1353  
 Db 3563 AGCGAGTGTGCAAAACCGGGGACCTGCGTGAACCTGTCTCATTTGAGGCTTCCACTGT 3622  
 QY 1354 AspCysProSerGlyAspPheGluYrProYrCysGlnValThrArgSerPhePro 1373  
 Db 3623 GGTGTCCCGCCCGCGAGATATGACATCCCTACTGTGAATGACACAGACCTTCCCA 3682  
 QY 1374 AlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeu 1393  
 Db 3683 CCCAGTCTCTGTGTTACTTCCAGAGGCTGCGGCAACGCTTCCACTTCCCTCTCCCTG 3742  
 QY 1394 SerPheAlaThrLeuGluArgAspGlyLeuLeuLeuYrAsnGlyYrAspPheAsnGlyCys 1413  
 Db 3743 GCGTTTGCCACCGAGGACAGAAATGCGCTGTCTCTACATGAGCCGCTTCATGAGAG 3802  
 QY 1414 HisAspPheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPheSerAlaGly 1433  
 Db 3803 CAGACTTATGCGCTGTGAGATTTGTGAGAGAGAGCTGCACTCACTTCTCGAGAGT 3862  
 QY 1434 GluSerThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrHis 1453  
 Db 3863 GAGACCAACAACCGGTGACACCGGAGTCTCTGAGGTGTGAGCGATGGCGGTGGCAT 3922  
 QY 1454 ThrValGlnLeuYrYrThrAsnYrProLeuLeuGlyGlnThrGlyLeuProGlnGly 1473  
 Db 3923 TCGGTCTGTGTGAGTACTACAAACAGCCCAATTTGGCACTGTGGGCTTCCCGCCAGG 3982  
 QY 1474 ProSerGluGlnLeuValAlaValAlaValThrValAspGlyCysAspThrGlyValAlaLeu 1493  
 Db 3983 CGGTGTGAGAGAGAGTGTGTGTGTGTGATGATGATGATGATGATGATGATGATGATG 4042  
 QY 1494 ArgPheGlySerValLeuGlyAsnYrYrSerCysAlaAlaGlnGlyYrThrGlnGlySer 1513  
 Db 4043 CACTTTGAGAAATTAAGTGTGGGAACTACAGCTCGCTGCGCCAGGCACTGAGCGGCTCC 4102  
 QY 1514 LeuLeuSerLeuAspLeuThrGlyProLeuLeuGluGlyValProAspLeuProGlu 1533  
 Db 4103 AAGATCTCATGTGATGTGATGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4162  
 QY 1534 SerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSer 1553  
 Db 4163 GACTTCCCGGTGACAGCGGTGAGTTGTGGAGTGCATGCGAAACCTGTCCATCGATGAGC 4222  
 QY 1554 ArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrValProGlyCysProAla 1573  
 Db 4223 CGGATTTGTGACATGTGCTCGTTTATTTGCAACAAATGTATCAAGGCGAGGCTGTGCTTCT 4282  
 QY 1574 LeuLeuAsnValCysAspSerAsnThrCysHisAsnGlyGlyYrThrCysValAsnGlnTrp 1593  
 Db 4283 CAGAGAACTTCTGCGATGTGAGCTCATGTCAGAAAGGGGGACCTGTGTGAACAGTGG 4342  
 QY 1594 AspAlaPheSerCysGluCysProLeuGlyPheGlyGlyYrSerCysAlaGlnGlnMet 1613  
 Db 4343 AACACGTAATTATGTAGTGCCTGCGCTGCTTGTGGGAAAGAACTGTGAACAAGCTATG 4402  
 QY 1614 AlaAsnProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuPro 1633  
 Db 4403 CCAACCTTCAAGGCTTCACTGTGTGAGACGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4462  
 QY 1634 IleSerGlnProTrpYrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeu 1653  
 Db 4463 ATCTGTGTGCTTGTGATCTGTGGGCTCATGTTCCGGAACCGGAAAGAGATGTGTGCTG 4522  
 QY 1654 LeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisVal 1673  
 Db 4523 ATGAAAGCAACAGCTGCAAGTCTTCCAGGCTCCATCTCCAGATTTCTCAACAGCTCATCTC 4582  
 QY 1674 MetLeuSerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArg 1693



OY	2396	PhePheLeuThrLeuLeuAaArgIIeLeuAaSerXenGlnHISgLYIIeAaGAspLeu	2415
Db	6791	GTCTCTCTCTGCGCTGCTTCGAGACATGCGCTCTCAACTCGACACAGATCCACAGAAACT	6850
OY	2416	ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIIeAaGlnAlaAspLeu	2435
Db	6851	ATCCACGCTCTGTTCTTCTTCCACAGCTCATCTTCAAGTGGCGGCATCAACAGACTGAGAC	6910
OY	2436	ProPheAlaCySThrValIIleAlaGIIeLeuLeuHisPheLeuTyrlLeuCySThrPheSer	2455
Db	6911	CCGTTTCTCTGACAGAGTGTGCGCACTCTCTGACATACGTCTTCAGAGGCACTTGCC	6970
OY	2456	TrpAlaLeuLeuGlnAlaLeuHisIleuTyrlaArgAlaLeuThrGluValaArgAspValaAsn	2475
Db	6971	TGACACCTTGTTGGAGAACTTGACATGTCTACCGCATGCTGACAGAAAGTGGCAATGTGAC	7030
OY	2476	ThrGlyProMetAaGpHeTyrlTyrMetLeuGlyTyrlGlyValProAlaPheIIerThGly	2495
Db	7031	ACTGGGCGCAAGGGGCTTCTTACACAGTGTGGCTGGGGCAATCCGTCATGTGTACACGGA	7090
OY	2496	LeuAlaValaGlyLeuAspProGlnGlyTyrlGlyAaAspProAspPheCySTrpLeuSerIle	2515
Db	7091	CTGCGTGTGGCTTGACCTTCAGGGCGATGGAACCTTGACTTCTGTGCTGTCTCTT	7150
OY	2516	TyrAspThrLeuIIeTrpSerPheAlaGlyProValaIlePheAlaValSerMetSerVal	2535
Db	7151	CAGATATACCTGATTTGGAGCTTTGCTGGCGCTGTCCGGAACGGTTATATATCATCAACA	7210
OY	2536	PheLeuTyrlIleLeuAlaAlaArgAlaSerCyAlaAlaGlnaGlnGlyPheGluIys	2555
Db	7211	GTCATCTTTGGCTGTCTGCAAAAGGTTTCTCGCAAAAGAAACCATTTATTAAGAAAGA	7270
OY	2556	LysGlyProValaSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThr	2575
Db	7271	AAAGGGGGTGTCTCCATGCTAGAGACGGCTTCTCTGCTGCTGCTGTCTGTCTACACGCCAC	7330
OY	2576	TrpLeuLeuAlaLeuLeuSerValaAspSerThrLeuLeuPheHISTyrlLeuPheAla	2595
Db	7331	TGGCTGTGGGACTGTGGCGGTCAACAGTACACTTGTACTTCACTTACCTCTTGGCT	7390
OY	2596	ThrCybaenCybIIeGlnGlyProPheIIePheLeuSerTyrlValaValLeuSerIysGlu	2615
Db	7391	GCTTTCAGCTCTTCGACGGGCACTTTCCTCTCTGTCTTCTACGCGTGGCCACAGGAG	7450
OY	2616	ValaArgIysAlaLeuIys--LeuAlaCySerAaArgIysProSerProAspProAlaLeu	2634
Db	7451	GTGCGGAAAGCACTGAGGCGGTGTCTGTGGCAAGGAAGCTGACGTGATGACTCGGCC	7510
OY	2635	ThrThrIysSerThr--LeuThrSerSerTyrlaAspCybProSerProTyrlAlaAspGly	2653
Db	7511	ACCACTCGGGCACTCTGCTAACCGGCTCTCTCAACTGCAACAACACTTACAGGAAAGG	7570
OY	2654	ArgLeuTyrlGlnProTyrl-----GlyAaSerAlaGlySerLeuHisIleSerThSerArg	2671
Db	7571	TCCAGACATGCTCCGCGACCGCGCCGTGGGAGTCCACAGCTCTCTCGAACAAGTACACAGG	7630
OY	2672	SerGlyIysSerGlnProSerTyrlIleProPheLeuLeuArgIgluSerAlaLeuAsn	2691
Db	7631	GATGAAGGGGTCCAG-----AAATCAAGTGTCTCC	7660
OY	2692	ProGlyGlnGlyProProGlyIleuGlyIaAspProGlySerLeuPheLeu--GluGlyGln	2710
Db	7661	TCTGGCCCAAGCCCGTGAACCATAGGAACAACAATGATCTCTTCAATCCCTAGGAATCC	7720
OY	2711	AspGlnGlnHISAspProAspThrAaSerAaAspSerAspLeuSerLeuGluAspAspGln	2730
Db	7721	AAAAAAGCTCAAGCGCCCTGACTCTGACTCTGACAGTGAAGCTGTCCG--GACGACAC	7777
OY	2731	SerGlySerTyrlAlaSerThrHisSerSerAaSerGlnGluGluGluGluGluGluGlu	2750
Db	7778	AGTAACTCTTACGGCTCTTCAACAACAATCGACACGGAGGATGATGGCGGAGAGCTGAA	7837

Oy		2751	GluGluIuAlaIaPheProGlyGluGlnGlyTrpAspSerLeuLeuGlyProGlyValaGlu	2770
Db		7638	GACAAA-----TGGAATCCGGCTGGGGCCCCGC-----	7667
Oy		2771	ArgLeuProLeuHisSerThrProLys--AspGlyGlyProGlyLyValaPro	2789
Db		7668	-----CATACACCCCAGAAAGCAGATGCTGTGCCAACCAACGATCCCGCTGGGC	7915
Oy		2790	TyrProGlyAsp-----PheGlyThrThralAlaGluSerSerGlyAanglyAlaPro	2807
Db		7916	TGGCCCAAGACGAGCCTGGCTGGAGTGAACAATAAGAATTGGACTGA-GCCCCACT	7974
Oy		2808	GluGluArgLeuArgGluAsn-----Gly	2815
Db		7975	GAAAGTCAGACCAAAGTCAGAGTCGAGTTACACCGCAGGCGCAGGGCAATCACTGGGC	8034
Oy		2816	AspAlaLeuSer-----ArgGluGlySerLeuGly---ProLeuProGlySerSerAla	2832
Db		8035	GACCGGCGCTCTGACCCGGAAGTGGGGTCTGTGCCAAGCCAGTGCGCTGTATTAAGCAGC	8094
Oy		2833	GlnPro-----HisLysGlyIleLeuLysLysLysCysLeu-----ProThrIle	2847
Db		8095	CAGCCCAAGAGCAGCAGCGGAAAGCATCTCTAAAAACAATCACCTACCCGCGGCATT	8154
Oy		2848	Ser-GluLys-----SerSerLeuLeuArgLeuProLeuGluGlnGlyCysThrGlySerS	2865
Db		8155	GCCAGACACACCACTGAAGTCCCGGCTGCCAGAGAAAGCTGCTGATGTGACAGACAGCCC	8214
Oy		2865	rArgLysSer---SerLaseSerGluGlySerArGLyLYrProProRo-----	2880
Db		8215	CACATCTCTCCGCAATCTCTCTTGGCTCTGGCGATGTGTCTCAGTCCACTGACTGTGT	8274
Oy		2881	-----ArgProProProArgInSerLeuGlnGluLnuLnuAnglyValMetPr	2897
Db		8275	CATTACCATCAAGACTCCGAGGAGGACCAGGCGTAGCATCTCATATGGGGTG----	8329
Oy		2897	oileAlMeSerLelueYalaglyThr	2906
Db		8330	----GCCATGATATGTACGACAGGGAGT	8353
<b>RESULT 13</b>				
ID	ABZ42586	standard; DNA; 11965 BP.		
AC	ABZ42586;			
XX				
DT	04-MAR-2003	(first entry)		
DE	Human CELSR3 nucleotide seq ID NO:655.			
XX				
KM	G protein-coupled receptor; GPCR; anti-genic peptide; gene therapy;			
KM	G protein-coupled receptor modulator; antibody; immune-related disease;			
KM	growth-related disease; cell regeneration-related disease; AIDS; cancer;			
KM	immunological-related cell proliferative disease; autoimmune disease;			
KM	Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;			
KM	osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;			
KM	graft versus host disease; Parkinson's disease; multiple sclerosis; pain;			
KM	psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;			
KM	mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;			
KM	hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;			
KM	ulcer; gene; ds.			
XX				
OS	Homo sapiens.			
PN	WO200261087-A2.			
XX				
PD	08-AUG-2002.			
XX				
PF	19-DEC-2001; 2001WO-US50107.			
XX				
PR	19-DEC-2000; 2000US-257144P.			
PA	(LIFE-) LIFESPAN BIOSCIENCES INC.			

XX Burner GC, Roush CL, Brown JP;  
XX WPI, 2003-046718/04.  
DR P-PSDB; ABP81741.  
XX

PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
PT (GPCR), useful for diagnosing and designing drugs for treating  
PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
PT cancer or autoimmune diseases

PS Disclosure: Fig 1, 523pp; English.

XX The present invention describes antigenic peptides (I) comprising:  
CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
CC acids. Also described: (1) an assay for the detection of a particular  
CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity  
CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
CC an antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related diseases, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorder, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. ABP42523 to ABP42659 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention.

XX Sequence 11965 BP; 2239 A; 3697 C; 3645 G; 2384 T; 0 other;

XX Alignment Scores:

Pred. No.: 0 Length: 11965  
Score: 7836.50 Matches: 1570  
Percent Similarity: 66.71% Conservative: 424  
Best Local Similarity: 52.53% Mismatches: 817  
Query Match: 50.41% Indels: 179  
DB: 25 Gaps: 48

US-09-916-849A-3 (1-2923) x ABP42586 (1-11965)

QY 28 ProLeuLeuGlyAspGln-----ValGlyProCysarg 38  
DB CCTCTTGGGGGGGAGACAGACCTTTGCAAGAGAGTATGTCACACAGGGGCTGTCT 759  
QY 39 SerLeuGlySerArgGlyArgGlySer----- 48  
DB 760 CAG---GGGTCCCGGGCTCGGGGAAACAGCTCCCTCCCTTCAAGACTTTTGATTCCGC 816  
QY 49 -----GlyAlaCysAlaProMetGlyTyrPleuCysProSerSerLaseran 64  
DB 817 ACCACGGTCCCAAGCGGGTCTCTCCACAGGAGAGCTGGAGACAGCTCCCAAGAG 876  
QY 65 LeuTyrPleuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84  
DB 877 ---TGGGCAACCGCGCTCTGTGGGAAATTATGGCAACAG-----GAGCAGAGGTC 927  
QY 85 -----ProHisHisAspGlyLeuArgValTyrCysPProGluSerGluAlaHis 100  
DB 928 AGGCGGAGAGACCGACGACAT-----CCGAGCAGAGAGAGAGACAG 966  
QY 101 IlePro-----LeuProProAlaProGluLeuTyrCys 110

DB 967 CCCCCGGGAGAACTGTTCTTCAGGGGCTCGGAGATCTGACCCGAGCTG--GATTCAACA 1025  
QY 111 ProTyrSerCysArgLeuLeuGlyYlIleGlyHisLeuSerProGluGlyLeuThr 130  
DB 1026 CCAAGCAGCGGAGAGACAGCTCTGATCAGATTCA---GACCCCGGAGCTCTGAC 1082  
QY 131 LeuProGluGluHisProCysAlaLeuValProArg-----LeuArgCysGlnSerCys 148  
DB 1083 GCTCCCGAGCGGCGCCCAAGCGCATGCGCTCCGGGGTCTCTTCCGCTGCGCTTC--- 1139  
QY 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLeu 161  
DB 1140 ---CTCCGAGCGCGCCCGGCGGCTCCCGGAGCTCCCGCGCTCTGAAACCGAG 1196  
QY 162 SerProGluGluSerLeuGlyYlYArgArgLysArgAsnValAsnThrAlaProGlnPhe 181  
DB 1197 AAAGTAACCTGGCGAACCGGGGACGCTTGTCTCCCGGCAACCGGACCGGAGTTT 1256  
QY 182 GlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProValAla 201  
DB 1257 CCGCAGTCAACTACACAGACCTGGTGGCGGAGATGAGCAGCAGCAGCGGCTGCTA 1316  
QY 202 SerLeuArgAlaIleAspProAspGluGlyGluAlaGlyValGluLeuGlyThrMetAsp 221  
DB 1317 GCGGTGTGCTCAGAGACCGGAGCGCGGAGGCGGCGGCTTACTCTGCTGCG 1376  
QY 222 AlaLeuPheAspSerArgSerArgSerArgGlnPheSerLeuAspProValThrGlyValAla 241  
DB 1377 GCATCTAGAACACACCGCTGCTGAGCTGTTCAGCATCAGACCGGAGCGGCTTATC 1436  
QY 242 ThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261  
DB 1437 GGTACGGGGGAGCTGAGACCGGAGAGATGAGACCTCACTGCTGAGCCGCG 1496  
QY 262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAsp 281  
DB 1497 CAGGACACGGGCTGCGCGCTCTCGGCAACACGATGAGCGGCTGACAGTACAGTAC 1556  
QY 282 ThrAsnAspHisAspProValPheGluGlnGlnGlyTyrLysGluSerLeuArgGluAsn 301  
DB 1557 CGCAACGACCACTGCGCGGCTTTTGAAGACGAGATACCGGAGACCTTCCGAGAT 1616  
QY 302 LeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsn 321  
DB 1617 GTGAGAGGAGGCTACCTATCTCAGCTGCGTGCCTCACTGACGCGGAGCGCCCTCAAC 1676  
QY 322 AlaAsnIleLeuTyrArgLeuLeu-----GluGlySerGlySerProSerGluVal 339  
DB 1677 GCAACCTGCGCTACCGCTTCTGTGGGCGCCGACGCTGCGCGCTGCGCGCC 1736  
QY 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
DB 1737 TTGAGATTAATCAACGCTCCGCGCTCATCAGCACACGCGCCAGAGTGACCGGAGAC 1796  
QY 360 ValGluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379  
DB 1797 ATGAAAGCTATGAGCTGTGTGTGAAGCCAGCAGGCGCAGAAACCGGGCGCGGC 1856  
QY 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPhe 399  
DB 1857 TGGGCACTGTGCGCGTACATTACTGTCTTAACAGAAACGAAATGCTCTCACTTC 1916  
QY 400 SerGluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeu 419  
DB 1917 AGCAGAAAGCGCTACGAGGCGAGGTGCGGAGATGTGGCGCCCAACAGTGCTGTG 1976  
QY 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSerIleMet 439  
DB 1977 CGCTCAACGCGCATGACCGGAGCAGAGACGCAAGATTTGGTCACTACATCATC 2036  
QY 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
DB 2037 AGTGGCAATAGCGGTGACACTTTCATGACAGCTTCACTGCGAGATTCAGGTGTG 2096

QY 460 SerProLeuAspArgIuThrThrIleuGluTyrThrLeuArgValArgAlaGlnAspGly 479  
 Db 2097 GCACCTCTGAGCTTCGAGGAGAGAGATGATGCCCTTGGGCATCAGAGGCCGACGATGCT 2156  
 QY 480 G1YArgProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499  
 Db 2157 GGGCCGGCCACCGGCTGTCACCAACACCGGGCTGGCCGACCTCCGAGGTGGACATCAAT 2216  
 QY 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValPro 519  
 Db 2217 GACACACATTCCTATTGTCGAGCAGCCCTTCCAGATTCTCTGGAAGAAATGCTCC 2276  
 QY 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaArg 539  
 Db 2277 TTGGGTCACACGATCATTCAAGCAGTGCATCAGACCATGGGAGAAATGCCAG 2336  
 QY 540 LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
 Db 2337 TTGGAGTACCTCCCTAACCTGGTGGCACCTGATCTCTTTGTGATAAACAGCGCAC 2396  
 QY 560 G1TyrIleSerValAlaIleAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 Db 2397 GGCTGGGATCTCTGTGAGTGGTCCCTGACCGTGAAGTCTGTGGAGCATTCCTTTGGT 2456  
 QY 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 Db 2457 GTGGAGGCTGAGACCATGGGTGACCCCACTCTGCTCAGCAGTGCACCGTGAAT 2516  
 QY 600 ValLeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
 Db 2517 GTGTGTGACCTTATGACAAATCGGCTGAGTTCACAAATGAAGGATGTCACATCAAGCATG 2576  
 QY 620 AsnGluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaValAspAspAla 639  
 Db 2577 AATGAGATCAGCTGTGGGACCAAGTGTGACGCTGACCGAGTGAACCGGATGACC 2636  
 QY 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 Db 2637 AACGTGCACATGACATCAGATCAGCGGCGAACACCGGAATGCTTTGGCATCAGC 2696  
 QY 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleGluGluArg 679  
 Db 2697 ACCGAGGGGGTGTGGTGTGGTGAATCTGTGCTGTGCTGCTGACATGCAAGAGAACG 2756  
 QY 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
 Db 2757 TACTTCAAGCTGATCACTGCATCTGACCGTCCCTTCAATGATCACTGATGTCAC 2816  
 QY 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrVal 719  
 Db 2817 ATCAACATCAAGATGACCAACATCATCGGCGGCTTTTCAAGATGCCCATCTACTGATG 2876  
 QY 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2877 AGGTGATATGAAGATCGGCCAATGGGTGACCAATAGTGTGATCAATGCTGTGATGAT 2936  
 QY 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2937 GACGTGGGTAGATGCTCGATCACTATCTCTGAGAGAACACCTGCCCGCATGTCGC 2996  
 QY 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 Db 2997 ATTGATCAAGATCAAGAGCATTAATCAAGGCCCATTTAGCTATGAGACCAAGGTG 3056  
 QY 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnPheSerAspThrThr 799  
 Db 3057 ACCATCACTGAGTGTGATCAAGCTCGGAGCAATGGCATCCACGAAGGAGACACTACT 3116  
 QY 800 TyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 3117 TATGTGAGGTGATGATCAATGATGATGATGACAAATGCTCCACAAATTTGTGGCTCCAC 3176

QY 820 TyrGlnGlySerValTyrGluAspValProPheThrSerValLeuGlnIleSerAla 839  
 Db 3177 TATACAGGGCTGTCTCTGAGAGATGCCCACTTTACAGAGTCTGTGAGATCTCAGCC 3236  
 QY 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAsp 859  
 Db 3237 ACTGACCGGAGATCATATCCCAATGCGCGGTCCAGTACATTTCCAGAAATGTGAGAT 3296  
 QY 860 G1YAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeu 879  
 Db 3297 GGGAGATGAGATTTTACATTAAGCCCACTTGAATTTGTCGTACAGTAAAGCGGCTTA 3356  
 QY 880 AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspArgGlyMetPro 899  
 Db 3357 GACCGGAGAGCATGATCAAGTATAGATTTGATGCTGCTACGAGTGCAGAGATGTGCC 3416  
 QY 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
 Db 3417 CCACTCCGAGCTCCAGTCAATCCAGGTGATGTCAGAGATGTGAACGACATGCACT 3476  
 QY 920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 3477 GTCTTCCAGCTGAGGAGTTTGAAGTTCGGGTGAAGAAATGACATTTGGGCTCAGTG 3536  
 QY 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3537 GTGGCCAGATCATCGATGAGTGAACCTTGAAGGCCCCAATGATTAATGATACAG 3596  
 QY 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3597 ATCGTGAAGGGAGAACATCCCTGAGCTGTCCAAATGACATCTCTCTGAGAACCTGACG 3656  
 QY 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValValLeuValIleGlnAlaThr 999  
 Db 3657 GCACTCATTAACCTTACATATAGAGCTGCCCAAGATATGATATGTGTGGAGGCCA 3716  
 QY 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3717 TGTGCTCTTGTGTGATGAGCGGGCCACTGTGACAGTCCGCTGTGACAGATATGACAC 3776  
 QY 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
 Db 3777 AGCCGTGTCTAACACTTCAGATCTCTTCAACATATATGATCAACCGTTCAGAC 3836  
 QY 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 Db 3837 ACCTTCCGTCGGGCAATATTGGGCGCATCCAGATTAATGACCCGATGTCTCCAGCAC 3896  
 QY 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 Db 3897 CTCTTCTACTCTTTGAGCGGTGGCAATGAGCTGAGCTGTGTGATCAACAGACAGCT 3956  
 QY 1080 G1YGluLeuValLeuSerArgAlaLeuAspAsnAsnAspProLeuGluAlaIleMetSer 1099  
 Db 3957 GGGAGCTGCACTACGCGGAAGCTAGCAATTAACCGCCACTGATGCTCCATGTG 4016  
 QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db 4017 GTGACGTGACAGATGGCTGACAGCGTGAAGGGGAGTGTGTCTGTGGCGTGTGATC 4076  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 Db 4077 ATCAACGAGAGTGTGTGCCCAACAGCTGACCGTGGCTTTGAGAAACATTTGGCAGAG 4136  
 QY 1140 ArgPheLeuSerProLeuGluIleuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
 Db 4137 CGCTTCTCTACCGCTGTGGCGCGCTTCTCTCAAGGCGTGTGCTGCGTCTCCAG 4196  
 QY 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
 Db 4197 CCGCTGAGAGAGTCTTCAATCTTCAACATCCAGAACGACAGACAGAGTA---GGGGGCAAC 4253  
 QY 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyPro----- 1197

Db 4254 GGGCTAAATGATGTTCTCGGCGTAGCTCCAGTGGGCGGCGGCGCTGCAGGG 4313  
 Qy 1198 ProPheLeuProSerGluAapLeuGlnGluArgLeuTyrLeuAapSerLeuLeuThr 1217  
 Db 4314 CCTGGTTACAGCTCCGAGGAGCTGCAGAGACATTGACGTGGCGGCGGCGCTGGG 4373  
 Qy 1218 AlaLeuSerAlaGlnArgValLeuProPheAapAapAapAapAapAapAapAap 1237  
 Db 4374 GCTCGCTCCCTGCTCCAGCTGACCTGCTTCAGACAGACAGTGGCTGCAGAGGCTCTGT 4433  
 Qy 1238 GUAenTyrMetArgCysValSerValLeuArgPheAapSerSerAlaProPheIleAla 1257  
 Db 4434 GAGAACTAACATGAAATGCGGTGCTGCTCCGCTTGACTGCTGCGCGCTTCCTCGGCG 4493  
 Qy 1258 SerSerSerValLeuPheAapProIleAlaProValGlyLeuArgCysArgCysPro 1277  
 Db 4494 TGGGCTCCAGCTGTTCCGAGCCATCCAGCCATGCGCTGGCTGGCTGGCTGGCTGG 4553  
 Qy 1278 ProGlyPheThrGlyAapTyrCysGlnThrGlyValAapLeuCysTyrSerArgProCys 1297  
 Db 4554 CCCGATTCACCGGAGACTTTTGGCAGACCGAGCTGACCTGCTGCTACCTCAACCACTGT 4613  
 Qy 1298 GAlProHISGlyArgCysArgSerArgGlyGlyGlyTyrThrCysLeuCysAapArgGly 1317  
 Db 4614 CGCAACGCGGAGCTGCGCGCGCGCGAGAGAGCTACACGTCGCTGCGCGCGCGCG 4673  
 Qy 1318 TTTThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1337  
 Db 4674 TTCACCGGAG 4733  
 Qy 1338 LysAapGlyGlyTyrCysValAapLeuValGlyGlyPheLysCysAapCysProSer 1357  
 Db 4734 CCCAACGCGGAG 4793  
 Qy 1358 GlyAap---PheGlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSer 1376  
 Db 4794 GCGGCGCTGAGAGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4853  
 Qy 1377 PheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAla 1396  
 Db 4854 TTGTCATGTTTCCGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4913  
 Qy 1397 ThrLysGluArgAapGlyLeuLeuLeuTyrAapGlyArgPheAapGlnLysHisAapPhe 1416  
 Db 4914 ACAGTGCAG 4973  
 Qy 1417 ValAlaLeuGlnValIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1436  
 Db 4974 CTGGCCCTGGAACTGCTGGCTGGCCAGAGTGCAGCTCAATATTCACCGGTGAATCCAG 5033  
 Qy 1437 ThrThrValSerProPheValProGlyGlyValSerAapGlyGlnThrPheThrValGln 1456  
 Db 5034 ACCGAGTCAAGCCCAAGTTCAGAGGAGCTTGAGTGAAGAGAGAGAGAGAGAGAGAG 5093  
 Qy 1457 LeuLysTyrTyrAapLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGln 1476  
 Db 5094 CTGAGATTCACAAAG 5153  
 Qy 1477 GlnLysValAlaValValThrValAapGlyCysAapThrGlyValAlaLeuArgPheGly 1496  
 Db 5154 GACAAAGTGGCTGCTGAG 5213  
 Qy 1497 SerValLeuGlyAapLysSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLysSer 1516  
 Db 5214 GCTGAGATTCAGCACTACATGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5273  
 Qy 1517 LeuAapLeuThrGlyProLeuLeuGlyGlyValProAapLeuProGlnSerPhePro 1536  
 Db 5274 CTGAGACCTGACGAGGCTCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5333  
 Qy 1537 ValArgMetArgGlnPheValGlyCysMetArgAapLeuGlnValAapSerArgHisIle 1556

Db 5334 GTATCCCATTAAGACTTCATGCGGTGATGCGGAGCTGCACATTGATGCGCGGAGTGC 5393  
 Qy 1557 AapMetAlaAapPheIleAlaAapGlnGlyThrValProGlyCysProAlaLysLysAap 1576  
 Db 5394 GACATGCGGCTTTGTGTGCAATATATGCAACATGAGAGCTTGCACAGCTACAC 5453  
 Qy 1577 ValCysAapSerAapThrCysHisAapGlyGlyThrCysValAapGlnThrAapAlaPhe 1596  
 Db 5454 TTTTGTGACTAGGCGGCTTCGCAAGACAGTGGCTTCTGCTCGGAGAGCTGGGCGAGCTTC 5513  
 Qy 1597 SerCysGlnCysProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAapPro 1616  
 Db 5514 AGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5573  
 Qy 1617 GlnHisPheLeuGlySerSerLeuValAlaThrHis---GlyLysSerLeuProIleSer 1635  
 Db 5574 CACCATTTCCGAGGAG 5633  
 Qy 1636 GlnProTyrTyrLeuSerLeuMetPheArgThrArgGlnAlaAapGlyValLeuLeuGln 1655  
 Db 5634 GTGCGATGTAACCTGGGCTGGCATTTGCGACACGCGGACACGAGGGGGTCTGATGCA 5693  
 Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeu 1675  
 Db 5694 GTGACGCTGGGCGCACACAGAGCTCTTGGCAGCTAAGTCGGGGGTACTGTCTGTG 5753  
 Qy 1676 SerVal---GlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyArgAla 1694  
 Db 5754 ACAGTGCAG 5810  
 Qy 1695 AapAapGlyAapThrPheHisAlaGlnLeuAlaLeu---GlyAlaSerGlyGlyPro 1712  
 Db 5811 AGTATGCGCGGTGAGAGATGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5870  
 Qy 1713 GlnHisAlaIleLeu---SerPheAapTyrGlyGlnGlnArgAlaGlnGlyLysLeu 1730  
 Db 5871 GGCACACATGCTTATGCTGCTCACTGAGCTTACCTTCCAGAGACACATGGCGGTG 5930  
 Qy 1731 GlyProArgLeuHisGlyLeuHisLeuSerAapIleThrValGlyGlyLe---ProGly 1749  
 Db 5931 GGGAGTACGTCAG 5990  
 Qy 1750 ProAlaGlyGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAap 1769  
 Db 5991 AGTGCAG 6050  
 Qy 1770 ThrProGlnGlyValAapSer---LeuAapProSerHisGlyGlnSerIleAapValGln 1788  
 Db 6051 ACACCTGCTGCTCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6104  
 Qy 1789 GlnGlyCysSerLeuProAapProCysAapSerAapProCysProAlaAapSerTyrCys 1808  
 Db 6105 CTGAGCTGTGTGTGACCAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6164  
 Qy 1809 SerAapAapThrAapSerTyrSerCysSerCysAapProGlyTyrTyrGlyAapAapCys 1828  
 Db 6165 CGGAGCTCTGCGAGAGACTTTTCTTGTGACCTGCGAGAGAGAGAGAGAGAGAGAGAG 6224  
 Qy 1829 ThrAapValCysAapLeuAapProCysGlnHisGlnSerValCysThrArgLysProSer 1848  
 Db 6225 GTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6284  
 Qy 1849 AlaProHisGlyTyrThrCysGlnCysAapProAapThrLeuGlyProTyrCysGlnThr 1868  
 Db 6285 GCGCCCATGCTATACCTGTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6344  
 Qy 1869 ArgIleAapGlnProCysProArgGlyTyrThrGlyHisPheThrCysGlyProCysAap 1888  
 Db 6345 AGAGTGCAG 6404  
 Qy 1889 CysAapValSerLysGlyPheAapProAapProAapLysThrSerGlyGlyCysHisCys 1908  
 Db 6405 TGTGATTCACAAAG 6464



QY 1909 LygEluabnhiStrYrArgProGlySerProThrCyseuLeuCyAspCyTyPro 1928  
 Db 6465 AAGAGATTCCACATACGAGCGCGGGGAGTGAAGCTTGGCTCCCATGTGACTGTACCT 6524  
 QY 1929 ThGlySerLeuSerArgValCyAspProGluAspGlyGlnCyProGlyProGly 1948  
 Db 6525 GTGGGGTCCCACTCGGCTCATGTGACCCACAGCGGGGAGTGGCTGTCCCGACGA 6584  
 QY 1949 ValIleGlyArgGlnCyAspArgCyAspAsnProPheAlaGluValThrThrAsnGly 1968  
 Db 6585 GGCCTTGCGCCGAGTGCACAGCTGTGACGTCCTTCCGAGAGGAGACGCGAGGCG 6644  
 QY 1969 CyGluValAlaSerValAspSerCyProArgAlaIleGluAlaGlyIleTyrTrpArg 1988  
 Db 6645 TGCAGGAGTCTATGATGCTGCTCCCTAAGTCCCTGATGCTGTGTGTGGCCAG 6704  
 QY 1989 ThrArgPheGlyLeuProAlaAlaAlaProCyProGlySerPheGlyThrAlaVal 2008  
 Db 6705 ACAAGTTTGCGCTGCGCCACAGTGCCTGTCCCGGGGGCGCTGGGTGCTGTG 6764  
 QY 2009 ArgHisCyAspArgLysIleArgGlyTyrPheProAsnLeuPheAsnCyThrSerIle 2028  
 Db 6765 CGGCTGTGTATAGGCGCCAGGGTTGGCTGAGCGCCGACTTTCACACTTACTCCCT 6824  
 QY 2029 ThrPheSerGlyLeuGlyPheAlaGluArgLeuGlnArgAsnGlySerGlyLeuAsp 2048  
 Db 6825 GGCCTTTCAGAGCTCACTGTGCTGTGATGCTGATGCTGACGTAACAAGACGCGACTGAT 6884  
 QY 2049 SerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnIleStrAlaGly 2068  
 Db 6885 ACATAGGAGGCAAGAGCTGGCTCAGCGGCTACGGGAGGTGACTGGCCACTGTACAC 6944  
 QY 2069 TyrPheGlySerAspValIleValAlaIleTyrGlnLeuAlaThrArgLeuLeuAlaHisGlu 2088  
 Db 6945 TATTTTAGCCAAAGATTTCAGATCACTGCCGCTGTGTGGCCCACTGTGGCTTCCAG 7004  
 QY 2089 SerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnLeu 2108  
 Db 7005 AGGCATCAGAGGAGCTTGGGGCTGACAGCACACAGGATGCCCACTTCAATGAGAACTCG 7064  
 QY 2109 LeuArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisStrGlyLeuIle 2127  
 Db 7065 CTGTGGGCGGCTGTGACTGTGCTGCCACAGACACAGGGACTTGTGGCGGCGCTGGG 7124  
 QY 2128 GlnGlnThrGlnGly-----GlyThrAlaTrpLeuGlnIleStrYrGluAlaTyr 2144  
 Db 7125 CAGCGGGCCCTCGGGGCTCCCAAGAGCGCGGGACTGGTGAAGGCACTGGAGAGTAT 7184  
 QY 2145 AlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThr 2164  
 Db 7185 GCAAGCCACACTCGAGAGAAATATGAATCAATCAATCCATGAGGAGCTGTGTACG 7244  
 QY 2165 ProAsnIleValIleSerValAlaArgLeuAspLysGlyAsnPro-----AlaGlyAla 2182  
 Db 7245 CCTAATATCAGCTCAGCATTTGACGCAATGAGAGACCCCAAGTTCCTCCCGGGGGGCGCT 7304  
 QY 2183 LysLeuProArgTyrGluAla-----LeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
 Db 7305 CGCATCCTCGCTACCATAGAACCTTTTGAAGGCAAGTGTGGAGTCCACACACC 7364  
 QY 2201 ThrValIleLeuProGlySerValPheArgGlyThrProProValValAlaArgProAlaGly 2220  
 Db 7365 CATGTGCTGCTGCTCCCAAGTCCCAAGGCACTCCCATGTGAAAGTTCTGCCACAAAGC 7424  
 QY 2221 ProGlyGlnGlnGlnProGlnGlnLeuAlaArgArgGlnArgHisStrGlyLeu 2240  
 Db 7425 AGCAGCATAGAAATCCACACCTCAAGTGTGTCCCAACAGCCCGCCGAGAGCA 7484  
 QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 7485 GAGCTGGAGTCTCATATATCATCTCTCGTTTACGCACTTAAGGGGAGCTGTCCCT 7544

QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
 Db 7545 GGCAGATTCCAGAGAAAGCCGAGGTGCAAGCTTCTCCAGAACCCCGTCAATGAATCC 7604  
 QY 2281 ProValIleSerIleSerValHisAspArgGlyGlnLeuLeuProArgAlaLeuLys 2300  
 Db 7605 CCGGTGTGAGGTGTGTGTGTTCACAGACCAATTCCTAAGGGGAATCTGTGAGTCC 7664  
 QY 2301 ProValThrValGlnPheArgLeuGlnThrGlnGlnArgThrTyrPheProIleCyVal 2320  
 Db 7665 CCATCAGCTAGAGTTCCTGCTGTACAGACAGGAATGAGACAGGCAATCTGTGTG 7724  
 QY 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyTyrTrpSerAlaArgGlyCyGln 2340  
 Db 7725 CAGTGGAGCCCACTGGCTGGAGAGACATGGTGTGTGTGACAGACGGGACTGTGAG 7784  
 QY 2341 ValValPheArgAsnLysSerIleValSerCyGlnCyAsnHisMetThrSerPheAla 2360  
 Db 7785 CTGTGTACAGAAAGAGTCCCAAGCAAGTGTCTGTGAGCGGACAGGACCTTTGGG 7844  
 QY 2361 ValLeuMetAspValSerArgArgLys-----AsnGlyGlnIleLeuProLeuLysThr 2378  
 Db 7845 GTCTCATAGAGTCTCTCCCTGAGAGCTGAGCGGAGCACTGAGCTGTGTGTG 7904  
 QY 2379 LeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePheLeu 2398  
 Db 7905 TTCAACCAAGT 7964  
 QY 2399 ThrLeuLeuArgLysLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAla 2418  
 Db 7965 CTGAGCTTCGAGGCTCAAGTCAATGTGCTGGAGTCAATGCAATGTGCAAGCGGC 8024  
 QY 2419 LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAla 2438  
 Db 8025 CTGGGGGTGAGAGTCTCTCTCTGAGGATTCACAGAACCCAAATCAGCTGTG 8084  
 QY 2439 CyThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCyThrPheSerTrpAlaLeu 2458  
 Db 8085 TGCATGCAAGTGCATCTCTCTGACATCTTCTCTGACACCTTCCGAGGCTTTC 8144  
 QY 2459 LeuGlnAlaLeuHisLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyPro 2478  
 Db 8145 GTGCAGGGCTGACCTCTTACCGCATGCAAGTGTGAGCCAGCACTGTGACCGCGGC 8204  
 QY 2479 MetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaVal 2498  
 Db 8205 ATGCGCTTCTACATGCCCTGGGGCTGGGGCTGCCGTGCTGTGCTGTGCTGTG 8264  
 QY 2499 GlyLeuAspProGlnGlyTyrGlyAsnProAspPheCyStrPheSerIleTyrAspThr 2518  
 Db 8265 GGGCTGAGACCTGAGGGCTATGAGAACCTGTACTTCTGTGATTCAGTCCACAGGCC 8324  
 QY 2519 LeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyr 2538  
 Db 8325 CTCTATGAGCTTGTGTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8384  
 QY 2539 IleLeuAlaIleArgLysSerCyAla-----AlaGlnArgGlnGlyPheGlyLysGly 2557  
 Db 8385 CTCCTGCTGCGCGCACATCTGTCTCCAGAGGAGAGAGGAGCC-----AAGAAAGACC 8438  
 QY 2558 ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeu 2577  
 Db 8439 TCTGCACTAGCCCTTGTGAGCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 8498  
 QY 2578 LeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrTrpPheAlaThrCys 2597  
 Db 8499 TTGGGCTCTGAGAGTCAACACAGCATCTTACAGCTTCACTCATGCTGAGTCC 8558  
 QY 2598 AsnCyIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlnValArg 2617  
 Db 8559 TGGGCTCTCAGGGCTGCGGTGCTGTCTTCTGTGTCTTATATCAGATGCTCGG 8618  
 QY 2618 LysAlaLeuLysLeuAlaCys---SerArgLysProSerProAspProAlaLeuThrThr 2636



(e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle, ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke, ischaemic cerebrovascular disease, Alzheimer's disease or Pick's disease), disorders of vesicular transport (e.g. cystic fibrosis, diabetes mellitus, Grave's disease, or goitre), gastrointestinal disorders, disorders (e.g. ulcerative colitis, or gastric and duodenal disorders), autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic and protozoal infections. The NOVX proteins can be used as immunogens to produce antibodies and as vaccines. The NOVX nucleotide sequences may be used in chromosome mapping, identifying individuals from minute biological samples (tissue typing), and in forensic identification of a CC biological sample. The present sequence encodes human NOV2, which is CC located on chromosome 3q21.3-4.

XX Sequence 12348 BP; 2287 A; 3911 C; 3696 G; 2446 T; 8 other;

# Alignment Scores:

Pred. No.:	Length:	12348
Score:	7815.50	1566
Percent Similarity:	66.14%	Conservative: 428
Best Local Similarity:	51.94%	Mismatches: 819
Query Match:	50.28%	Indels: 203
DB:	24	Gaps: 47

US-09-916-849a-3 (1-2923) x ABQ82327 (1-12348)

```

Qy 28 ProleuLeuGlyAspGln-----ValGlyProGlyArg 38
Db 416 CCTCTTCGGGGCGGAGACGCTTTCGAAAGAGTACTGTCTACCGAGGGCTCTGTCT 475
Qy 39 SerLeuGlySerArgGlyArgGlySerSer----- 48
Db 476 CAG---GGGTCCCGGGCTCGGGAGAACAGCTCGCCCTCCCTCAGACTTTTATGTCGC 532
Qy 49 -----GlyAlaCysGlyAlaProMetGlyTyrLeuGlySerProSerSerAlaSerAsn 64
Db 533 ACCACGCTCCCAACCGCGGTCTCTCCAGCGGAAAGCTCGGAGACGCTCCCGCAAAAGAG 592
Qy 65 LeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84
Db 593 ---TGCGACCGCGCGCTCTCTGGGAAATTATGGCAACAG-----GAGCAAGGGCTC 643
Qy 85 -----ProHisAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100
Db 644 AGGCGAGAGAGCCACGACAT-----CCGAGCAGAAAGAGCAG 682
Qy 101 IlePro-----LeuProProAlaProGluGlyCys 110
Db 683 CCCCCCGGAGACTGTCTTCAGAGGCGCTCGGAGATCGGCCCGGAGCTG-GATTACAGA 741
Qy 111 ProTyrSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGluGlyLeuThr 130
Db 742 CCACGACACGGGAGAGACAGCTCTCTGCATCAGATTCA---GCACCCCCGAGCTCTCGACA 798
Qy 131 LeuProGluGluHisProCysLeuValAspAlaProArg-----LeuArgCysGlnSerCys 148
Db 799 GCTCCCGAGCGCGGCCCAAGCGCATGGCTCCCGGGGTCTCTTCGCTGCGCGCTTC--- 855
Qy 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyGluArg 161
Db 856 ---CTCCCGAGCGCGGCCCGCGCTCCCGCGGAGCTCCCGGCGCTCTGAGAGCAGG 912
Qy 162 SerProGluGluSerLeuGlyIleArgArgLysArgAsnValAsnThrAlaProGluPhe 181
Db 913 AAATGAACCTCGGGGAAACCGGGACGCTTCTGTGCGCCCGGCAACCGCCACCGCATTT 972
Qy 182 GlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProValAla 201
Db 973 CCGCAGTACACACTACGACCGCTGTGCGGAGATGAGCAGACGCGCGGTCTCTA 1032
Qy 202 SerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluTyrThrMetAsp 221
::: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1033 CGGTGGTTCCTCAGGACCCGAGCGCGCGGAGCGGCGGCTAGTCTACTCGTGGCG 1092
Qy 222 AlaLeuPheAspSerArgSerArgGlnPheSerLeuAspProValThrGlyAlaVal 241
Db 1093 GACTCATACACGCGCTCGCTGAGCTGTTCATGATGACCCGAGAGCGGCTTATC 1152
Qy 242 ThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261
Db 1153 CGTAGCGGCGGAGCTCTGAGACCGGAGAGCATGAGAGCTCATCACTCGGTGACCGCG 1212
Qy 262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAsp 281
Db 1213 CAGAACACAGGGTTCGCGCGCTCTCGGCACACAGATGTCGCGTGCAGTACGAGCGAC 1272
Qy 282 ThrAsnAspHisAspProValPheGluGlnGlnGlyTyrLysGluSerLeuArgGluAsn 301
Db 1273 CGCACGACCTCTCGCGGTTTTCAGCAAGCGGACGATCCGGAGACCTTTCGCGAGAT 1332
Qy 302 LeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProAsn 321
Db 1333 GTGAGAGAGGGCTACCTTATCTCGAGCTGCGTGCATGACGCGGACGCGCCCAAC 1392
Qy 322 AlaSerIleLeuTyrArgLeuLeu-----GluGlySerGlyGlySerProSerGluVal 339
Db 1393 GCCAACCTCGCTTACCGCTTCGTGGACCGCAGCTCGCGGCTCGACCTCCCGCGCC 1452
Qy 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlu 359
Db 1453 TTGAGATTGATTCACGCTCCGGCTCATCAGACCAAGGCGCGAGTGAACCGGAGAC 1512
Qy 360 ValGluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379
Db 1513 ATGAAAGCTATGAGGTGTGTGTGAGACGACGACAGAGGCGAGGAAACCGGCGCGCC 1572
Qy 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPhe 399
Db 1573 TCGGCACCTGCGCGGTACATCTGTCTGTAGAGAAACAGACATCTCTCCAGTTC 1632
Qy 400 SerGluLysArgTyrValValGlnValArgGluAspValThrProGlyValaProValLeu 419
Db 1633 AGCGAGAACCGCTTACGTCGCGAGAGTCCGAGAGTGTGCGGCCCAACACATCGTGTG 1692
Qy 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSerIleMet 439
Db 1693 CGGTGTCACGGCCTGACCGGAGCAAGGCAACGATGTGTGATCATCAACATCACC 1752
Qy 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459
Db 1753 AGTGGCATAGCCGTGAGACATTTGCTCATGACAGCTCATGCGAGATCCAGGTGTG 1812
Qy 460 SerProLeuAspTyrGluThrThrLysGluTyrThrLeuArgValArgAlaGlnAspGly 479
Db 1813 GCACCTCTGACCTTGAGCAGAGAGAGATGTGCTTCGATCAGCGCGCAGAGTGTCT 1872
Qy 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499
Db 1873 GGCCTGCGCACCCCTGTCCAAACAACGCGGCTCGGACGATCAGAGTGTGTGACATCAAT 1932
Qy 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519
Db 1933 GACACATTCCTATTATTTTGTACAGCGCCCTTCGAAAGTTTGTCTGTGAAATAAGTCTCC 1992
Qy 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539
Db 1993 TTGGGTCACTCAATCATCAATCAAGAGATGAGATGAGATGAGATGAGATGAGATGAG 2052
Qy 540 LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559
Db 2053 TTGGAGTACTCTCTTACCTGTGTGCACTGTATCTCTTGTGTATTAACAGCGCACT 2112
Qy 560 GlyTyrIleSerValAlaIleAspLeuAspArgGluGluValAlaAspPheTyrSerPheGly 579
Db 2113 GGTGGGTCTCTGTAGTGTGCTCCCTGACCGTGAAGTCTGTGAGATTAATCTTTGTGT 2172

```

QY	580	ValGIuValAaRgAspHisGIYThrProAlaLeuThrAlaSerAlaSerValSerValThr	599
Db	2173	GTGAGGCTCGAGACACATGGCTCACCCCACTCTGTGCTCAGCAAGTGCACCGTAGCT	2232
QY	600	ValIleuAspValAsnAspAsnAsnProThrPheThrGlnProGluTYrThrValArgLeu	619
Db	2233	GTGCTGACGTTAAATGACATCGGCTGAGTTCCAAATGAAGAGATACACCTACGACTG	2299
QY	620	AsnGIuAspAlaAlaValAlGIYThSerValValThrValSerAlaValAspArgAspAla	639
Db	2293	AATGAGATGACAGCTGTGGGCAACCATGTGTGCACCGCAGACCGCAGTGAACCGTAGTCC	2352
QY	640	HisSerValIleThrTYrGlnIleThrSerGIYAsnThrArgAsnArgPheSerIleThr	659
Db	2353	AACGTGCGATCAGCTACCATCCAGATCCAGCGGCACACCCGGAAATGCGCTTGCCATCAAGC	2412
QY	660	SerGlnSerGIYGIYGIYLeuValSerLeuAlaLeuProLeuAspTYrLYLeuGIuArg	679
Db	2413	ACCCAGGGGGGTGGTGGGTCTGGTGTACTGGCTGTGCATCGACATTCACAGAGAAAGCC	2472
QY	680	GlnTYrValLeuAlaValThrAlaSerAspGIYThrArgGlnAspThrAlaGlnIleVal	699
Db	2473	TACTTCAAGCTGGTACTAACTGCATCTACACGTCGCTTCATAGATCACTGTAATGTGAC	2532
QY	700	ValAsnAlaThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYrThrVal	719
Db	2533	ATCAACATCAAGATGCCACCACTCATCGCGCGCTTTCAAAGTGCCCATCATCTCAAGTG	2592
QY	720	AsnValAsnGIuAspArgProAlaGIYThrThrValValLeuIleSerAlaThrAspGIu	739
Db	2593	AGTGTGAATGAAAGATCGGCCAATGGGTGACACCATAGTGATCAACAGCTCTGATGAT	2652
QY	740	AspThrGIYGIuAsnAlaArgIleThrTYrPheMetGluAspSerIleProGlnPheArg	759
Db	2653	GACGTGGGTAGAAATGCTCTGATCACTCATCTCTGGAGAGACACCTGCCCAATTCCGC	2712
QY	760	IleAspAlaAspThrGIYAlaValThrThrGlnAlaGluAspTYrGIuAspGIuVal	779
Db	2713	ATTATGAGAGACTCAGAGCCATTACATTACAGGCCCATTTAGACTATGAGACCAAGGTG	2772
QY	780	SerTYrThrIleuAlaIleThrAlaArgAspAsnGIYIleProGlnIleSerAspThrThr	799
Db	2773	ACCTACACCTCGCTATATACAGCTCGGAGCAATGGCAATCCACAGAAAGGAGACATACT	2832
QY	800	TyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer	819
Db	2833	TATGTGAGGTGATGTGCTCAATGACGTGAATGACATGTCTCCAAATTGTGGCTCCCAAC	2892
QY	820	TyrGlnGIYSerValTYrGIuAspValProProPheThrSerValLeuGlnIleSerAla	839
Db	2893	TATACAGGCGTGTCTGTGAAGATGCCCACTTTCACAAATGTCTGCAATCTCAGCC	2952
QY	840	ThrAspArgAspSerGIYLeuAsnGIYArgValPheTYrThrPheGlnGIYAlaAspAsp	859
Db	2953	ACTGACCGGAGATGCTCAATGCCAATGGCCGGGTCCAGTACATTTCCAGAAATGGTGAAGT	3012
QY	860	GIYAspGIYAspPheIleValGIuSerThrSerGIYIleValArgThrLeuArgArgLeu	879
Db	3013	GGGATGTGAATTTTACATTTGAGGCCCACTCTGGAATGTCCGTACAGTAAAGCGGCTCTA	3072
QY	880	AspArgGIuAsnValAlaGlnTYrValLeuArgAlaTYrAlaValAspLYArgIlePro	899
Db	3073	GACCGGAGGAGCATATCAGTATGAATTTGACTGTCTTACGCAATGGAGACAGAGGTGCTCC	3132
QY	900	ProAlaArgThrProMetGluValAlaThrValThrValLeuAspValAsnAspAsnPro	919
Db	3133	CCACTCCGGAATCCAGTCAATCCAGGTATGATGGTGCAGATGTGAACAGCAATGACCT	3192
QY	920	ValIleGlnGIuAspGIuPheAspValPheValGIuGluIleuAsnSerProIleGIYLeuAla	939
Db	3193	GTCTTCCAGCTGAGAGATTTGAAGTGTGGGTGAAAGAAATACATTTGTGGTGTCAAGT	3252

QY	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetThrGln	959
Db	3253	GTGGCCCAAGATCATCTGACGTAGGAACTCTGACAGAAAGCCCAATGCCATATATATGACAG	3312
QY	960	IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluMetThr	979
Db	3313	ATCTGGAGGGGGAACATCTCCCTGACCTGTCCAAATGAGACATCTTCTGTGAGAACTGACG	3372
QY	980	AlaIleuValAspLeuAspArgGlyuAspArgProGluArgValIleGluAlaThr	999
Db	3373	GCACCTACATGACCTAGACTATGAGAGCTCTGCCAAAGAAATATGTAATGTGCTGACGCCCA	3432
QY	1000	SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAsn	1019
Db	3433	TCTGCTCTTTGGTTCAGCGCGGCGACACTGTGACCTCCGCTGGTTCACAAATATGACAC	3493
QY	1020	ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSer	1039
Db	3493	AGCCCTGGCTGCACCAACTTCAGATCCCTTCAACAACATATGATATCAACCCGTTCAGAC	3552
QY	1040	SerPheProGlyGlyAlaIleGlyValArgValProAlaHisAspProAspIleSerAspSer	1059
Db	3553	ACCTTCCCGTCGGGCATTTATTTGGCGCATCTCCAGCTTATGACCCCAATGTCTCCGACAC	3612
QY	1060	LeuThrTyrSerPheGluArgGlyAsnGluIleuSerLeuValLeuMetAsnAlaSerThr	1079
Db	3613	CTCTTCTACCTCTTTGAGCGTGGCAATGACCTGACGCTGGTATGATCAACCAAGCAGT	3672
QY	1080	GlyGluIleuLeuLeuSerArgAlaIleuAspAsnArgProLeuGluAlaIleMetSer	1099
Db	3673	GGGAGAGCTGCAGCTACCGCCAAAGCTAGACATTAACGCCCACTGGCGCTTCATGTG	3732
QY	1100	ValIleuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle	1119
Db	3733	GTGACTGTGCACAGATGCGCTGCACAGGTGACCGCGCAGTGTGCTGCCGTGTATC	3792
QY	1120	IleThrArgGluMetLeuThrHisAspIleThrLeuArgLeuGluAspMetSerProGlu	1139
Db	3793	ATCACGGAGAGATGTGCTGGCCAAACAGCTACCGGTGGCCCTTGAGAACATGTGCAGAGAG	3852
QY	1140	ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr	1159
Db	3853	CGCTTCTGTGCACCGCTGCTGGCGCGCTTCTCTGAGGGCGGTGGCTGGCTGCTGCATACG	3912
QY	1160	ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis	1179
Db	3913	CCCGCTAGAGACGCTTCATCTTCAATCATCAATCAACAACACACAGACGTA---GGGGGCACG	3969
QY	1180	IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyPro-----	1197
Db	3970	GTGCTCATATGATGATTTCTTGGCGCTAGCTGCACAGTGGGGCCGGGGCGGCGCTGCAGGG	4029
QY	1198	ProPheLeuProSerGlyuAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThr	1217
Db	4030	CCCGTGTTCAGCTCCGAGAGAGCTGCAGAGACAGATTGATCGTGGCGGGCGGGCGCTGGCG	4089
QY	1218	AlaIleSerArgAlaIleArgValLeuProPheAspAspAsnIleCysLeuArgGluProCys	1237
Db	4090	GCTCGCTCCCTGTGCTGCACGTACCTGCCCTTGCAGAGACACGTGGCCCTGCAGAGGCCCTGT	4149
QY	1238	GlyuAsnTyrMetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleAla	1257
Db	4150	GAGACACTAATGAATGATCGTGTCTCGTCTCCGCTTTGACCTGCTCCGGCCCTTCTGGCC	4209
QY	1258	SerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysPro	1277
Db	4210	TGGGCCCTGCACGCTGTTCGACCACTCCACCTCCAGCTGAGCTGGCGCTGCCCTGCCCTGCCG	4269
QY	1278	ProGlyPheThrGlyAspArgTyrCysGluThrGluValAspLeuCysArgTyrSerArgProCys	1297
Db	4270	CCGGATTCACGGGAGACTTTTCTCGAGACCCAGACTGACCTTGCTACTCTCAACCCATGT	4329
QY	1298	GlyProHisGlyValArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGly	1317

```

Db      4330 CGGACGGCGGAGCGCTGCGCGCGCGGAGGAGGCTACACGTCGCGTCCCGCGCGC 4389
Qy      1318 TyThrGlyValHisCysGluValIserAlaArgSerGlyValCysThrProGlyValAlaCys 1337
Db      4390 TTCACCGGAGAGAGACTGCGAGCTGAGACCGAGGCGCGCGCTGCGCGCGCGGCGTCTGC 4449
Qy      1338 LysAsnGlyValThrCysValAsnLeuValGlyValPheValCysAspCysProSer 1357
Db      4450 CGGACGCGGGGCACTGACCGGACCGGCCCAACGGCGGCTTTCGCTGCGCGCGCGCA 4509
Qy      1358 GlyAsp---PheGlyValProGlyCysGlnValThrArgSerPheProAlaHisSer 1376
Db      4510 GGGGCGCGCTTCGAGGGCGCGCGCTGCGAGGTGCTGCGCGCTTCCTCCGCGCGCAGTTCCG 4569
Qy      1377 PheIleThrPheArgGlyLeuValArgGlnArgPheHisPheThrLeuAlaLeuSerPheAla 1396
Db      4570 TTCGTCATGTTTCGCGCGCTGCGCGCGAGATTCACCTTACGCTTCCCTCTGTTGCGG 4629
Qy      1397 ThrLysGluArgAspGlyLeuLeuLeuValAsnGlyValArgPheAsnGlyValHisAspPhe 1416
Db      4630 ACAAGTCAGAGAGCGGGCTGCTCTTCTTCAACAGCGGCGCTGAACGAGAGCAAGCACTTC 4689
Qy      1417 ValAlaLeuGluValIleGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThr 1436
Db      4690 CTGGCCCTGAGACTCGTGGCTGGCCAGTGGCGGCTCAATATTCACGCGGTGAATCCAC 4749
Qy      1437 ThrThrValSerProPheValProGlyValValSerAspGlyValIleThrValGln 1456
Db      4750 ACCGTGTCAGCGCCCAACAGTTCCAGGGGCTTGAAGTACCGGCAATGGCATACAGTGCAT 4809
Qy      1457 LeuValSerValAsnValProLeuLeuGlyValIleThrValLeuProGlnGlyProSerGlu 1476
Db      4810 CTGAGATACCTACCAAGCCCGGACAGATGCCCTAGGGGGTGCAAGCGGCGCTCCAG 4869
Qy      1477 GlnLysValAlaValValThrValAspGlyCysAspPheGlyValAlaAlaLeuArgPheGly 1496
Db      4870 GACAGGTGCTGCTGACAGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 4929
Qy      1497 SerValLeuGlyValAsnValSerCysValAlaGlnGlyValThrGlnGlyValSerValSer 1516
Db      4930 GCGGAGATTGGCACTACCTATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4989
Qy      1517 LeuAspLeuThrGlyProLeuLeuLeuGlyValValProAspLeuProGlySerPhePro 1536
Db      4990 CTGACCTGACCGGCGCTCTCTCTCTCTGAGAGTCTCCCAACCTCCCGGAGAACTTCCCC 5049
Qy      1537 ValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIle 1556
Db      5050 GTATCCCATAGGACTTCATCGGCTGATGCGGGACCTGCACTGATGCGCGCGCAAGTG 5109
Qy      1557 AspMetAlaAspPheIleAlaAsnAsnGlyValThrValProGlyCysProAlaValSerVal 1576
Db      5110 GACATGCGGCGCTTTCGCAAAATATGAGCACATGCGAGGCTGCCAAGCAAGCTACAC 5169
Qy      1577 ValCysAspSerAsnThrCysHisAsnGlyValThrCysValAsnGlnIleProAspAlaPhe 1596
Db      5170 TTTTGTGACTCAGCGCGCTGCAAGAACAGTGGCTTCTGCTGCGAGCGCTCGGGCAGGCTTC 5229
Qy      1597 SerCysGluCysProLeuGlyPheGlyValSerCysValGlnGlnMetAlaAsnPro 1616
Db      5230 AGCTGCACTGCGCTGAGGCTTGGCGGCAAGACTGTCAGCTTACTTATGGCCCAATCCC 5289
Qy      1617 GlnHisPheLeuGlySerSerLeuValAlaIleThrHis---GlyLeuSerLeuProIleSer 1635
Db      5290 CACCATTTCCGTTGCAACGGGACACTGAGCTGAACTTTGGAGTGCATAGGCGTGTCT 5349
Qy      1636 GlnProThrValLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGln 1655
Db      5350 GTGCGCATGAGTACCTGGGCTGGGCACTTTCGAGCACGAGCNAACGAGGGGGTCTCGATGCA 5409
Qy      1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeu 1675

```

```

Db      5410 GTGAGGCTGGGCAACAYNNACGCTCTTTCGACAGTACATGCGGGGTACTGCTGTG 5469
Qy      1676 SerVal---GlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAla 1694
Db      5470 ACAATGACCAAGGGGCTCGGGC---CGTCTTCCATCTCTTCTGACCAAGGTACTGTC 5526
Qy      1695 AsnAspGlyAspTrpHisIleAlaGlnLeuAlaLeu-----GlyAlaSerGlyPro 1712
Db      5527 AGTATGCGCGGTGACACATCTCGGCTGAGGTTCAGAGAGAACACAGTGGCGCGG 5586
Qy      1713 GlyHisAlaIleLeu-----SerPheAspValGlnGlnArgAlaGlnValAsnLeu 1730
Db      5587 GGCCACCATGCTTATAGTCTCACTGACTTACCTTTCACAGACACCATGCGCGTG 5646
Qy      1731 GlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyValIle---ProGly 1749
Db      5647 GGGAGTGAAGCTCAGGGGCTGAAGATGAAGCAAGCTCAAGTGGAGGCTGCGCGCGC 5706
Qy      1750 ProAlaGlyValAlaAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAsp 1769
Db      5707 AGTGCAGAGAGAGCTCTCAGGGGTCTGTTGGCTCATTCAGAGGGGTGTGCTCGGCTCC 5766
Qy      1770 ThrProGlnGlyValAsnSer---LeuAspProSerHisGlyGlySerIleAsnValGlu 1788
Db      5767 ACACTCTGCTGCTCCCGCGCTGCTACCCCAAGCTAC---GAGTGAATCGGAG 5820
Qy      1789 GlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerValCys 1808
Db      5821 CCGCTGTGTGTGATCCAAACCCCTGTGCTGTGGGCGCTGCGCCACCTCAGCAGACATGC 5880
Qy      1809 SerAsnAspTrpAspSerValSerCysSerCysAspProGlyValValValAspAsnVal 1828
Db      5881 CGGACCTCTGGCAACCTTTCTTTCGACCTGCGAGCCAGGTATCAAGCGCCAGAGCTGT 5940
Qy      1829 ThrAsnValCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLysProSer 1848
Db      5941 GTGATGCTGCTGCTCTTAACCTCTGTCAACAGAGATATACCGGAGCTGCCAGCA 6000
Qy      1849 AlaProHisGlyValThrCysGlnCysProProAsnValLeuGlyProValCysGlnThr 1868
Db      6001 GCGCCCAATGACTATACCTGTGATCTGTGAGGTGCTATTTGGGACACATGTCAGGAC 6060
Qy      1865 ArgIleAspGlnProCysProArgGlyValTrpTrpGlyHisProThrCysGlyValProCysAsn 1888
Db      6061 AGATATGACCAAGCAAGTGCACAGGGGCTGTGTGGGAGACCCCAACTGTGGCGCTGCAAC 6120
Qy      1889 CysAspValSerValGlyPheAspProAspCysAsnValThrSerGlyValCysHisCys 1908
Db      6121 TGTGATGTTCAAAAGGTTTATTCCTCAACTGACACAAAGCAATAGGCAAGTGTACTGC 6180
Qy      1909 LysGluAsnHisValArgProProGlySerProThrCysLeuLeuCysAspCysValPro 1928
Db      6181 AAGAGATTCACATACCGACCGGGGGCAGTACTCTTGCCTCCATGTGATCACTACCT 6240
Qy      1929 ThrGlySerLeuSerValArgValCysAspProGluAspGlyValGlnCysProCysValProGly 1948
Db      6241 GTGGGTCTCACTCGCGCTCATGTGACACCCCAAGCGGAGATGCGCTGTGCGCCAGGA 6300
Qy      1949 ValIleGlyValArgGlyCysAspArgCysAspAsnProPheAlaGlyValThrArgHisGly 1968
Db      6301 GCCCTTGGCGCGCACTGCAACAGCTGTGACAGTCCCTTTCGAGAGTGCACACGCGC 6360
Qy      1969 CysGluValAsnValThrAspSerCysProArgAlaIleGlnAlaGlyValLeuTrpTrpProArg 1988
Db      6361 TCCCGGCTGCTGATGATGCTGCGCTTAAGTCTCCGATGATGTGTGTGTGTGGCCCGG 6420
Qy      1989 ThrArgPheGlyValProAlaAlaAlaProCysProLysGlySerPheGlyValThrValAla 2008
Db      6421 ACAAGTTTGTGGCTCGCGCCACAGTGCCTGTCCCGGGGGGCGCTGTGATCTGTG 6480
Qy      2009 ArgHisCysAspAspGlnHisIleArgGlyValTrpLeuProProAsnLeuPheAsnValThrSerIle 2028
Db      6481 CGGCTGTGTGATGAGGCGCAGGCTTGTGAGCGCCGACCTTTCACTGTACTCTCCCT 6540

```

Qy	2029	ThrPheSerGlnIleuValGlyPheAlaGlnArgLeuGlnArgAlaGlnIleuSerGlyLeuAsp	2048
Db	6541	GCCCTTTCGAGAGCTCTCAGCTCTGCTGCTGAGGCTTACAGTGAACAGACAGACCTGAT	6600
Qy	2049	SerGlyArgSerGlnGlnIleuAlaLeuLeuLeuArgAlaIaThrGlnIleThrAlaGly	2068
Db	6601	ACCATGAGAGCCCAAGAGAGCTGGCTTCAGGGGCTACGGAGAGTGACTGGCCACACTGACAC	6666
Qy	2069	TyrPheGlySerAspValLysValAlaTyrGlnIleuAlaThrArgLeuIleuAlaIleGlu	2088
Db	6661	TATTTTATGCCAAAGATGTTCCAGATCACTGCGCCGCTGCGAGCCCACTGCTGGCTTCGAG	6720
Qy	2089	SerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlnAsnIleu	2108
Db	6721	AGCCATACAGCAGGGGCTTCGGGCTGACACCCACACAGAGATGCCCACTTAATGAGAAATGTG	6780
Qy	2109	LeuArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisTyrGlnIleuIle---	2127
Db	6781	CTGTGGGGCCGGCTCTGCACTGCTTGGCCCAAGACAGGGAGACTTGTGGGGCCGGCTGGGG	6840
Qy	2128	GlnGlnThrGlnGly-----GlyThrAlaTyrLeuLeuGlnHisTyrGlnAlaTyr	2144
Db	6841	CAGGGGCCCCCTGGGGGGCTCCCAAGGCGCGGGGACTGTGTGAGGCACTGAGAGAGTAT	6900
Qy	2145	AlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThr	2164
Db	6901	GCACCCACACTCGCAAGGAATATGAGACTCACTACATTCCTGAATCCCATGGGGCTGGTGAAG	6960
Qy	2165	ProAsnIleValIleSerValValArgLeuAspLysGlyAsnPhe-----AlaGlyAla	2182
Db	6961	CCTAATATCATGCTCAGCATTTGACCGGATGAGGACCCCAAGTCTCCCGGGGGGCCGT	7020
Qy	2183	LysLeuProAlaGlyTyrGlnAla-----LeuArgGlyGlnGlnProProAspLeuGlnThr	2200
Db	7021	CGTACCTCGCTACCATATGACAACTCTTGTGAGGCGAGATGCTGGGAATCTCAACACC	7080
Qy	2201	ThrValIleLeuProGlnLysSerValPheArgGlnThrProProValValArgProAlaGly	2220
Db	7081	CATGTGCTGCTGCTTCCCACTCCCAAGGCGCAATCCCATCTGAAGTTCTGCGCCACAAAG	7140
Qy	2221	ProGlyAlaGlnGlnProGlnGlnIleuAlaArgArgGlnArgGlnAspGlnIleu	2240
Db	7141	AGCAGCATAGAAACTCCACCACTCAATGTGTGTCCTCCCAACAGCCCGCCAGAGCCA	7200
Qy	2241	SerGlnGlyGlnAlaValAlaSerValIleIleTyrArgThrIleuAlaGlyLeuLeuPro	2266
Db	7201	GAGCTGGGAGATCTCATATATCATCTCTCGTTTACCGCACTTAAGGGGACTGCTCCT	7266
Qy	2261	HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr	2286
Db	7261	GCCCAAGTTCAGGAGAAAGCCGAGAGTCCAGGCTTCTCAAGAACCCGTCATGAACTCC	7320
Qy	2281	ProValValSerLysSerValHisAspAspGlnIleuLeuProAspAlaLeuAspLys	2300
Db	7321	CCGGTGTGTACGGTGTGTGTCCACGAGCGCAACTTCTTAAGGGGAATCTGGAGTCC	7386
Qy	2301	ProValThrValGlnPheArgLeuLeuGlnThrGlnAlaGlnTyrPheProIleCysVal	2320
Db	7381	CCCATCAGCCCTAAGATTCGCTCTGTACACAGGAATCGAAGACAGGCAATCTGTGTG	7440
Qy	2321	PheThrAsnHisSerLysLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGlyCysGln	2346
Db	7441	CAGTGGGACCACTGGGCTGGCGGAGACATGTGTGTGACAGACACGGAGACTGGAG	7500
Qy	2341	ValValPheArgAsnLysSerHisValSerCysGlnCysAsnHisMetThrSerPheAla	2366
Db	7501	CTGTGTGCACAGAAATGGGTCCCAACCAAGTGTCTGCTCACCCGGAACGGGACTTGTGG	7566
Qy	2361	ValLeuMetAspValSerArgArgLys-----AsnGlyGlnIleLeuProLeuIleThr	2378
Db	7561	GTCCTCATGATGCTCTTCCCGGTGAGAGGCTGAGGGGCACTGAGAGCTGTGCTGTG	7620

OY	2379	LeuThrYrValAlaIleuGlyValThreLeuAlaIleuLeuLeuThrPhePheLeu	2398
Db	7621	TTTACCACAGCTGGCTGGCTGGCTGTCTGTGGCTGCGCTGGCTGACTGACGCCATCTGT	7680
OY	2399	ThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAla	2418
Db	7681	CTGAGCCTCGCGAGCCTCAAGTCCAAATGTGCGTGGATTCATGATGCCAATGTGGCAGCGCC	7740
OY	2419	LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGln	2432
Db	7741	CTGGGGGTGGAGACTCTCTCTCTCTCTCTGGGATTCACAGAACCCAAATCAGGTGCAG	7800
OY	2432	-----	2432
Db	7801	GATCAGGCGCAGGGAACTTGATGCTCATATGACCTACTGGCGCCAGAGAGCCTGGGGCGAA	7866
OY	2433	---AlaAspLeuProPheAlaCysGlnValIleAlaIleLeuLeuHisPheLeuYrLeu	2451
Db	7861	AATCCAGAGGTTCAGAGCTGGTGTGTGACATGACAGTGCATCTCCCTGCACTACTCTTCTCTC	7920
OY	2452	CysThrPheSerThrAlaLeuLeuGlnAlaLeuHisIleLeuYrArgAlaLeuThrGluVal	2471
Db	7921	AGCACCTTCCGCTGGCTCTTCTGTGACAGGGCTGTGACCTCTTCCGCATGCAAGTTGAAGCA	7980
OY	2472	ArgAspValAsnThrGlyProMetArgPheYrThrLeuGlyYrGlyValProAla	2493
Db	7981	CGCAACGTGACCGCGCGGCATGCGCTTCAACATGCCCTGGGCTGGGGCGTCCGTGCT	8040
OY	2492	PheIleThrIleLeuAlaValGlyLeuAspProGlnGlyYrGlyAsnProAspPheCys	2511
Db	8041	GTCGTGCTGGGCTCTGTGTGGGTGTGACCCCTGAGGGCTATGAGAACCCCTGACTTCTGC	8100
OY	2512	TrpLeuSerIleYrAspThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaVal	2531
Db	8101	TGATCTCAAGTCCACGAGCCCTCATGTGAGAGCTTGTCTGGCCTGTATGTCCTGGTCATA	8160
OY	2532	SerMetSerValPheLeuYrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGln	2551
Db	8161	GTCATGACCGGACCAAGTTCCTCTGTCGCGCAGCACTCTGCTCCACAGGCGAGAG	8220
OY	2552	GlyPheGlnIleYrGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeu	2571
Db	8221	GAGGCCAAGAAAGACCTGTGACTCAGGACCTTGCAGCTCTCTTCGCTGCTCTGCTG	8280
OY	2572	LeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHis	2591
Db	8281	GTCAGTGCCTTCGCTCTTCTTGGGCTCTCTGGCACTCAACACAGCATCTTACCTTCCAC	8340
OY	2592	ThrLeuPheAlaThrCysAsnGlyIleGlnGlyProPheIlePheLeuSerYrValVal	2611
Db	8341	TACTCTCAGTGTGAGCTCTGCGGCTTCAGAGGCTGGCGGGTGTGCTGCTCTTGTGTCTC	8400
OY	2612	LeuSerIleGlnValArgIleAlaIleLeuYrLeuAlaCys---SerArgIleProSerPro	2630
Db	8401	CTAATATGCAGATGCTGGGTGTGCTGTATGCAGCCTGTCTGGGCAAGAAAGCAGCGCT	8460
OY	2631	AspProAlaLeuThrThrIleYrSerThrLeuThrSerSerYrIleAsnYrProSerProYr	2650
Db	8461	GAGAGGCAAGGCCAGACACTGGGTGTGGAGCTGGGGCTCTAACACACAGCGCTCTTCTT	8520
OY	2651	AlaAspGlyArgLeuYrGln---ProYrGlyYrAspSerAlaGlySerLeuHisSerThr	2669
Db	8521	GAGAGAGTGGCTCATCCGCATCATCTGTGGGCGCTTCACCGCTCTCTCTGTAGACACT	8580
OY	2670	SerArgSerGlyIleYrSerGlnPro-----SerYrIle-----	2680
Db	8581	GCCGCGCTCCGCGCGAGACCAAGACACAGACAGCCAGCGGCGCGACACTCTCAGAGAC	8640
OY	2681	ProPheLeuLeuArgGlnGlnSerAla-----LeuAsnProGly	2693
Db	8641	AATGTCTGTGTGACATGAGCTCAGCCGCTGACCACTGATCCACAGCCTTCAGGCTCAT	8700
OY	2694	GlnGlyProProGlyLeuGly-----AspProGlySerLeuPheLeu	2707





DB: 22 Gaps: 49  
 US-09-916-849A-3 (1-2923) x AAD08065 (1-11648)

28 ProLeuLeuGlyAAspGln-----ValGlyProCysArg 38  
 416 CCTCTTGGCGGCGGACAGGACCTTTGCAAGAGTAGTGTGTGACAGAGGGCTGTCTCT 475

39 SerLeuGlySerArgGlyArgGlySerSer-----48  
 476 CAG-----GGGTCCGGGGCTGGGGAGACAGCTCCGCCCTTCAGACTTTTATTCGGC 532

49 -----GlyAlaCysAlaProMetGlyTyrLeuCysProSerSerAlaSerAsn 64  
 533 ACCACGGATCCCAAGCCGGGTCTCTCCAGCGGAAACCTGGGACAGGCTCCCGCAAAAGAG 592

65 LeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84  
 593 ---TGGGCAACCGGCGCTGTGTGGGGAATTATGGCAACAG-----GGAGCAAGGGTC 643

85 -----ProHisIAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100  
 644 AGGGCGAGAGGCAAGACAT-----CCGAGCAAGAAAGACAG 682

101 IlePro-----LeuProProAlaProGluGlyCys 110  
 683 CCCCCCGCGGAACTGTCTTCAGAGGGCTTCGGGATCTGGCCCGAGCTG-GATTCAAGA 741

111 ProTyrSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGlnIleGlyLeuThr 130  
 742 CACGCGACCGGAGGACAGCTCTGCATCAGTTCA--GCACCCGCGCAAGTCTCGGACA 798

131 LeuProGluGluHisProCysLeuValAlaProArg-----LeuArgCysGlnSerCys 148  
 799 GGTCCCGAGCCGCGCGCCCAAGCGCAAGCCCTCCCGGGTCTCTCCGCTGCGCTTC--- 855

149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLeuArg 161  
 856 ---CTCCCGCAGCGCCCGCGCGCGCTCCCGCGAGCTCCCGCGCTCTTAAGCAAG 912

162 SerProGluGluSerLeuGlyIleArgArgLysArgAsnValAsnThrAlaProGlnPhe 181  
 913 AAAGTAACCTCGGCAACCGGGGACGCTTTCGTGCGCGCAACCGGCAACCGGCAAGTT 972

182 GlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyTyrProValAla 201  
 973 CCGGAGTAACTACACAGCGCTGCGCGGAGAAAGAGGACGAGGACCGCGGTGCTA 1032

202 SerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluTyrThrMetAsp 221  
 1033 CGCGTGGTGTCTAGAGCCCGGACCGCGCGGAGCGCGCGCTAGTCTACTCGCTGGCG 1092

222 AlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAlaVal 241  
 1093 GCACTCATAGACAGCGCTGCGCTGAGCTGTACAGTACACCGGAGCGCGCTTATC 1152

242 ThrThrAlaGluGlyLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261  
 1153 CGTACGGCGGCGGAGCTGTGACCGCGAGACAGATGAGCGTCACTACTGCTGTGACCGG 1212

262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAsp 281  
 1213 CAGGACACAGGGTGTGGCGGCTCTCGGCACACAGATGTGGCGGTGACATACCGGAC 1272

282 ThrAsnAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGluAsn 301  
 1273 CGCAACAGACACTCGCGGCTTTTAAAGCAAGCGGACATCCGGAAGCCCTTGGCAAGAT 1332

302 LeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsn 321  
 1333 GTGGAGAGAGGCTACCCCTTCTCGAGCGTGGCGGACATGACGGGAGCGCGCCCAAC 1392

322 AlaAsnIleLeuTyrArgLeuLeu-----GluGlySerGlySerProSerGluVal 339

DB 1393 GCCAACCTCGGCTACCGGCTTGTGGGGCGCGGACGCTCGCGGCTACGCTCCGCGCC 1452  
 GY 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
 DB 1453 TTCAGATTGATCCAGCGCTCCGCGCTCATACACACAGCGCGCGAGTGGACCGGAGCAC 1512

GY 360 ValGluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379  
 DB 1513 ATGGAAGCTATGACTGTGTGTGAGACCGGACCGAGGCGAGAACCCCGGCGGCC 1572

GY 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAlaProGlnPhe 399  
 DB 1573 TCGGCACACTGTCGCGGTACATACATCTGTGTAGAGCAAGACATATGCTCTCAAGTTT 1632

GY 400 SerGluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeu 419  
 DB 1633 AGCGAAGCGCTACGTGCGCGAGGTGCGCGAGATGTGCGCCGCCACACAGTGTGTG 1692

GY 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValHisTyrSerIleMet 439  
 DB 1693 CGGTACGCGCACTGACCGGACACGAGACCGCAACGGATTTGTGCTACATCAATCATC 1752

GY 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
 DB 1753 AGTGCATATGACCGGTGACACTTTGCCATGCAGACGCTCATGTGCGAGATCCAGGTGTG 1812

GY 460 SerProLeuAspTyrGluThrThrLysGluTyrThrLeuArgValArgAlaGlnAspGly 479  
 DB 1813 GCACCTGTGACTTCGAGCGAGAGAGATGTGCTTGCCTGACAGGCGCGAGATCT 1872

GY 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499  
 DB 1873 GCGCGGCAACCGCTGTCCAAACACGSGCTGGCAGCATCCAGTGTGACATCAAT 1932

GY 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
 DB 1933 GACCAATTCCTAATTGTCAGACGCGCTTCCAAGTTTCTGTGTGAAATAATCTCC 1992

GY 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 DB 1993 TTGGGTCACTAGTCATCAATTCAGGACGTCAATGCAACATGGGAGAAATGCCAGA 2052

GY 540 LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
 DB 2053 TTGGAGTACTCCCTAACGTGTGTGACACTGATCTCTTTTGTGATTAACGCCACT 2112

GY 560 GlyTyrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 DB 2113 GGTGTGGTCTGTGTGAGTGTGCCCTTGACCGGTAGTGTGTGAGCATTACTTCTTGT 2172

GY 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrIleAspAlaSerValSerValThr 599  
 DB 2173 GTGGAGCTCGAGACCATGTGCTCACCCCACTGTGTCTAGGCAAGTACCGTGACT 2232

GY 600 ValLeuAspValAlaAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
 DB 2233 GTGTGACAGTTAATGACAATCGCGCTGAGTTCAATGAGAGATACACTACGACTG 2292

GY 620 AsnGluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaValAspArgAspAla 639  
 DB 2293 AATGAGATGACGTGTGGGACCAAGTGTGTGACCGGAGCGGAGTGAACCGGTATGCC 2352

GY HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThr 659  
 DB 2353 AACAGTGCATCACTACATACAGATCAACAGCGGCAACCCGGAATCGCTTGCATACG 2412

GY 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluLysArg 679  
 DB 2413 ACCGAGAGGAGGTGTGGGTGTGTGACTGTGCTGTGCACTGAGCTACAGAGGAAGCC 2472

GY 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699

Db 2473 TACTTCAGCTGGTACTAAGTCACTGACCGGCTTCATGATCACTGCTATGTGCAC 2532  
 Qy ValAsnVal ThrAspAlaAsnThrHisArgProVal PheGlnSerSerHisGlyThrVal 719  
 Db 2533 ATCAACATCAAGATGCGCACTCATCGCGGCTTTCAAAATGCGCCACTACTACGTG 2592  
 Qy AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2593 AGTGTGATGAAATGCGCCCAATGGGTAGCACTAAGTGTGCATCACTGCTCTGATGAT 2652  
 Qy AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2653 GACGTGGGTGAAATGCTCGTATCACTATCTCTGGAGGACCACTCCGCCCGCTTCGCG 2712  
 Qy IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 Db 2713 ATTGATGACAGCTAGAGAGCCATTACATACAGCCCATTAAGCTATGAGACCAAGTGTG 2772  
 Qy SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
 Db 2773 ACCTACACCTCGCTATCAAGCTCGGGGACAAATGGCATCCACAGAGGACAGACTACT 2832  
 Qy TyrLeuGluIleLeuValAspAspValAspAspAsnAlaProGlnPheLeuAspSer 819  
 Db 2833 TATGTGAGGTGATGTCATATGACGTAAATGACAAATGCTCCACAAATTTGGCTCCAC 2892  
 Qy TyrGlnIleSerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 2893 TATACAGGGCTGCTCTGAGAGATGCCCACTTTCACAGTGTCTGCAAGATCTCAGCC 2952  
 Qy ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnIleGlyAspAsp 859  
 Db 2953 ACTGACCGGAGTGTCTCATGCGCAATGCGGGGTCAATGACATTTCCAGAAATGGTGAAT 3012  
 Qy GluAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuAspArgLeu 879  
 Db 3013 GGGGATGAAATTTACATGAGCCCACTCTGAAATGTCTGTAACAGTAAAGCGGCTA 3072  
 Qy AspArgGluAsnValAlaGlnTyrValLeuAspArgIleTyrAlaValAspArgGlyMetPro 899  
 Db 3073 GACCGGAGGACGATGACGTGATGAGTTCGCTACGAGTTCGACAGAGGTGTGCC 3132  
 Qy ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
 Db 3133 CCACTCGGAGCTCCAGATGATCCAGATGATGATGATGATGATGATGATGATGATGAT 3192  
 Qy ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 3193 GTCTTCCAGCTGAGAGTTTGAAGTGGGTGAAAGATGATGATGATGATGATGATGATG 3252  
 Qy ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3253 GTGGCCAGATCACTGACGTGAGCCCTGACGAGAGGCCCAATGCGCATTAATATATACAG 3312  
 Qy IleValGluGluAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3313 ATCTGTGAGGGGAACTCCCTGAGCTGTTCCAATGACATCTTCTCTGAGAACTACAG 3372  
 Qy AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 3373 GCATCTATGACTAGACTATGAGGCTGCGCAAAATATGATGATGATGATGATGATGATG 3432  
 Qy SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3433 TCTCTCTCTTGGTCAAGCGGGCCACTGTGCACTGCGCTGTTGACAGAAATGACAC 3492  
 Qy ProProValIleuGluAsnPheGluIleuPheAsnAsnTyrValThrAsnArgSerSer 1039  
 Db 3493 AGCCTGTGCTCAACACTTCAAGTCTTCAACAACATATGATCAACCGTTTCAGAC 3552  
 Qy SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAlaIleSerAspSer 1059  
 Db 3553 ACCTTCCGTGGGCAATTATGGGCGCATCCCACTTATAGACCCGATGTCTCCGACAC 3612

Qy LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1060  
 Db 3613 CTCTTTACTCTCTTGAAGGTGGCAATGAGCTGCACTGCTGTGATCAACAGACCAAT 3672  
 Qy GluGluLeuLeuSerArgAlaLeuAspAsnAspArgProLeuGluAlaIleMetSer 1099  
 Db 3673 GGGGACCTCGACTAGCGCGGAAGCTAGACATTAACCGCCACTGGTGGCTTCATGTTG 3732  
 Qy ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuAspValThrIle 1119  
 Db 3733 GTGACTGTACAGATGAGCTGTGACAGCTGACGCGGCACTGTGTGTGCGGTGTACAC 3792  
 Qy IleThrAspGluMetThrHisSerIleThrLeuArgGluAspMetSerProGlu 1139  
 Db 3793 ATCAACGAGAGATGTGTGCGCAACAGCTGACCGTGTGCTGAGAACTATGGCAGAG 3852  
 Qy ArgPheLeuSerProLeuGluGlyLeuPheIleGlnAlaValAlaThrLeuAlaThr 1159  
 Db 3853 CGCTTCTGTCAACGCTGTGAGGCGCTTCGAGAGCGGTGGCTGCGCTGCTACG 3912  
 Qy ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlnHis 1179  
 Db 3913 CCGCTGAGAGCTCTTCACTTCAACATCCAAACAGACAGAGCTA--GAGGACACC 3969  
 Qy IleLeuAsnValSerLeuSerValGlyGlnProGlyProGlyGlyPro----- 1197  
 Db 3970 GTGCTCAATGTAGATTTCTCGCGCTAGCTCAACGTGGGCGCGGCGGCGCTGACAGG 4029  
 Qy ProPheLeuProSerGluAspLeuGlnIleArgLeuTyrLeuAsnArgSerLeuLeuThr 1217  
 Db 4030 CCTGGTTCAGCTCCAGAGAGCTGCAAGAGAGTTTATGCTGCGCGGCGCGCTGAGG 4089  
 Qy AlaIleSerAlaGlnArgValLeuProPheAspAsnAlaCysLeuAspGluProCys 1237  
 Db 4090 GCTCGCTCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4149  
 Qy GluAsnTyrMetArgCysValSerValLeuAspArgPheAspSerSerAlaProPheIleAla 1257  
 Db 4150 GAGAACTATGATGAATAGCTGCTCGGCTCCGCTTGACTGCTGCGCGCTTCTGCGCC 4209  
 Qy SerSerSerValLeuPheArgProIleHisProValGlyGlyLeuAspArgCysPro 1277  
 Db 4210 TCGGCTTCAACCTGTTCGACCCATCCAGCCATCCGCTGCGCTGCGCTGCGCTGCGC 4269  
 Qy ProGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCys 1297  
 Db 4270 CCGGATTCACGGAGACTTTTGGACAGCCAGCTGCACTGCTGCTGCTGCTGCTGCTG 4329  
 Qy GluProIleGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGly 1317  
 Db 4330 CGGACGGGAGAGCTGCGCGCGCGCGGAGGAGGCTACACGTCGCTGCGCGCGCGCGC 4389  
 Qy TyrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCys 1337  
 Db 4390 TTACCGGAGAGAGACTGACGAGCTGACACCGAGCGCGCGCTGCTGCGCGCGCTGCTG 4449  
 Qy LysAsnGlyGlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSer 1357  
 Db 4450 CGGACGGGAGAGCTGACCGAGCGCGCGCGCGCGCGCGCTTTCGCGCGCGCGCGCGCA 4509  
 Qy GlyAsp--PheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSer 1376  
 Db 4510 GCGGCGCGCTGCGAGCGCGCGCGCTGCGAGGTGTGCGCGCTGCTTCCCGCGCACTTG 4569  
 Qy PheIleThrPheArgGlyLeuArgProGlnArgPheHisPheThrLeuLeuLeuSerPheAla 1396  
 Db 4570 TTGATATGTTTGGCGCTGCGCGCGAGCAATTCATTAAGCTGTGCTTCTGCTTGGCG 4629  
 Qy ThrGluGluArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluHisAspPhe 1416  
 Db 4630 ACGGTGACAGAGAGCGGCTGCTTCTTCAACGCGCGCTGAGACGAGAGACGACCTTC 4689

QY 1417 ValAlaLeuGluValIleGlnGluValGlnLeuThrPheSerAlaGlyGlnSerThr 1436  
 Db 4690 CTGGCCCTGGAACTCTGGCTGGCCAGAGCGGCTCAATATTCACGGGTGAATCCAAC 4749  
 QY 1437 ThrThrValSerProPheValProGlyGlyValaSerAspGlyGlnThrPheThrValGln 1456  
 Db 4750 ACCGGGTAGAGCCCAACAGATCCAGGGGGCTTGAGAGCGGGCAATGGCAATGACAGTCAT 4809  
 QY 1457 LeuLeuValTyrAsnValSerProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGln 1476  
 Db 4810 CTGAGATTAACAACAAGCCCGGACAGATGCCCTTAGGGGGTGCACAGGCCCTCCCAAG 4869  
 QY 1477 GlnLeuValAlaValaThrValaAspGlyCysAspThrGlyValAlaLeuArgPheGly 1496  
 Db 4870 GACAAGTGGCTGTGTGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 4929  
 QY 1497 SerValaLeuGlyAsnValSerCysAlaAlaGlnGlyThrGlnGlyGlySerValSer 1516  
 Db 4930 GCTGAGATTTGGCACTACTACATGCGGCTGCTGGTGCACAAACAGCTCCAAAGATCC 4989  
 QY 1517 LeuAspLeuThrGlyProLeuLeuGlyGlyValaProAspLeuProGlnSerPhePro 1536  
 Db 4990 CTGGAACCTGAGCGGCTCTTCTTCTGGAAGGTGCTCCCACTCCCGAAGAACTTCCCC 5049  
 QY 1537 ValaArgMetArgGlnPheValaGlyCysMetArgMetLeuGlnValaAspSerArgHisIle 1556  
 Db 5050 GATATCCCAATGAAGACTTCATCGGCTGTATGCGGAGACTGCACATGATGATGCGCGAGT 5109  
 QY 1557 AspMetAlaAspPheIleAlaAsnGlnValaThrValaProGlyCysProAlaLeuValSer 1576  
 Db 5110 GACAATGGCGGCTTTGTGTCACAAATATGACACATGACAGCGCTGCACAAACAGCTAC 5169  
 QY 1577 ValCysAspSerAsnThrCysHisAsnGlyGlyThrCysValaAsnGlnThrAspAlaPhe 1596  
 Db 5170 TTTTGTGACTAGGCGGCTCGAAGAAACAGTGGCTTCTGCTCGGAGCGCTGGGCGAGCTTC 5229  
 QY 1597 SerCysGluCysProLeuGlyPheGlyGlyCysValaGlnGluMetAlaAsnPro 1616  
 Db 5230 AGCTCGACATGCGCTGTGGCTTGGCGGCAAGCTGACGCTTACATGATGCGCCATCCC 5289  
 QY 1617 GlnHisPheLeuGlySerSerLeuValaIleThrPheGlyLeuSerLeuProIleSer 1635  
 Db 5290 CACCATTTCCCTGGGCAACCGGCACTGAGCTGAACCTTTGGAAGTGAACATGCTGTGCT 5349  
 QY 1636 GlnProTyrTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValaLeuLeuGln 1655  
 Db 5350 GTGCCATGGTACTCGGGGCTGGCATTTGGGACACGGGCAACGAGGGGCTCTGATGCA 5409  
 QY 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlyHisIleValMetLeu 1675  
 Db 5410 GTGCAGGCTGGGCAACAGCGCTCTTGGCTGAGTGGGGGTGATGCTGTCTGTG 5469  
 QY 1676 SerValaGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAla 1694  
 Db 5470 ACAATGACCAAGGGGCTGGGCG--CGTGGCTTCCCATCTCTTGTGACACAGGTGACTGTC 5526  
 QY 1695 AsnAspGlyAspTyrPheHisIleAlaGlnLeuAlaLeuGlnAlaSerGlyPro 1712  
 Db 5527 AGTATGGCGGCTGGCAGATCTGGGGCTGGAGTTGGCAGAGAAACAGGTGGCGGGGG 5586  
 QY 1713 GlnHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValaLeu 1730  
 Db 5587 GGCCACCAATGCTTATGCTCATCTGACCTTACCTTCTTCACAGACCAATGCGCGT 5646  
 QY 1731 GlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValaGlyGlyProGly 1749  
 Db 5647 GGAATGAGCTGACAGGCTGAGAGTGAAGCAAGCTCCACATGGAGGCTGGCGCGGCG 5706  
 QY 1750 ProAlaGlyGlyValaAlaArgGlyPheArgGlyCysLeuGlnGlyValaArgValaSer 1769  
 Db 5707 AGTGACAGAGAGGCTCTCAGGGTCTGGTTGGCTGATCCAG----- 5748  
 QY 1770 ThrProGlnGlyValaAsnSerLeuAspProSerHisGlyGlnSerIleAsnValaGln 1789

Db 5749 -----CCACCGAGT-----GAATCGGAGCT 5769  
 QY 1790 GlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSer 1809  
 Db 5770 GAGTGTGTGGTACCAACGCTGTGCTTGGGCTTCCCACTACAGCAACATGCGCG 5829  
 QY 1810 AsnAspTyrAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyValaAsnValSerThr 1829  
 Db 5830 GACCTTGGCAGACCTTTCTTGTGACCTGACAGCGAGTTACTAGGCCAGGCTGTG 5889  
 QY 1830 AsnValCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLeuProSerAla 1849  
 Db 5890 GATGCTCTCCCTGTAACCCCTGTGACAGAGGATCATCGGCACTGCAAGAGCC 5949  
 QY 1850 ProHisGlyTyrThrCysGluCysProProAsnTyrLeuGlyProTyrCysGlnThrArg 1869  
 Db 5950 CCCATGGCTATACCTGTGACTGTGTGGTGGATTTTCGGGACACATGTGAGCAAG 6009  
 QY 1870 IleAspGlnProCysProArgGlyTyrThrPglHisProThrCysGlyProCysAsnCys 1889  
 Db 6010 ATGACACAGAGTGGCCACGCGGCTGTGTGGGAGGCCAACCTGTGCGCTCGCACTGT 6069  
 QY 1890 AspValSerTyrGlyPheAspProAspCysAsnValSerGlyGlyCysHisCysValys 1909  
 Db 6070 GATGTTCAAAAGGTTTGTATCCCACTGCAACAAGCAATGGGCAGTGTCACTGCAAG 6129  
 QY 1910 GlnAsnHisTyrTyrArgProProGlySerProThrCysLeuLeuCysAspCysTyrProThr 1929  
 Db 6130 GAGTTCCACTACCAACCGGGGAGTGAATCTTGGCTCCCATGTGATGCTTACCTGTG 6189  
 QY 1930 GlySerLeuSerArgValCysAspProGlyLeuAspGlyGlnCysProCysValProGlyVal 1949  
 Db 6190 GGCTCAACCTGGCTCATGTGCAACCCACAGCGGAGATGCCCTGTGCGCCAGAGACC 6249  
 QY 1950 IleGlyArgGlnCysAspArgCysAspAsnProPheAlaGlnValaThrThrAsnGlyCys 1969  
 Db 6250 CTGGCGCCGACAGTCAACAGCTGTGACAGTCCCTTCCGACAGAGTGAACAGCGGCTGCG 6309  
 QY 1970 GlnValaAsnTyrAspSerCysProArgAlaIleGlnAlaGlyIleThrPheProArgThr 1989  
 Db 6310 CGGGTCTGTGATGAGCTGTGCTTAAGTCCCTGAGATGTGTGTGTGTGTGTGTGTGTGT 6369  
 QY 1990 ArgPheGlyLeuProAlaAlaAlaProCysProGlySerPhe----- 2004  
 Db 6370 AAGTTGGCGTCCGGGCAACAGTGCCTGTGCTCCGCGGGGCGCTGGAGATTGGGGGTGCA 6429  
 QY 2005 GlnThrAlaValaArgHisCysAspGlnHisArgGlyTyrLeuProProAsnLeuPheAsn 2024  
 Db 6430 GGTGCTGTGTGGCGGTGTGTGATGAGGCCCAAGGGTTGGCTGAGACCCGACCTTTCAC 6489  
 QY 2025 CysThrSerIleThrPheSerGlnLeuLeuSerGlyPheAlaGlnArgLeuArgGln 2044  
 Db 6490 TGTACTCTCCCTGCTTTCAGAGCTCATGTGCTGCTGCTGAGTGGCTTGAAGCTGAACAAG 6549  
 QY 2045 SerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgValaIleThrGln 2064  
 Db 6550 ACCGACATGATTAACATGAGAGCCAGAGAGCTGTCTAGCGGCTTACAGGAGATGACTGCG 6609  
 QY 2065 HisThrAlaGlyTyrPheGlySerAspValaValaIleTyrGlnLeuAlaIleThrArgLeu 2084  
 Db 6610 CACACTGACCACTTTTATGACCAAGATGTTGAGTCACTGCGCCGCTGCGGCCCACTG 6669  
 QY 2085 LeuAlaHisGlnSerThrGlnArgGlyPheGlyLeuSerAlaIleThrGlnAspValHisPhe 2104  
 Db 6670 CTGGCTTTCAGAGCACTACAGAGGCTTGGGCTGTGACAGCAACAGAGATGCCCACTTC 6729  
 QY 2105 ThrGlnLeuLeuLeuArgValaGlySerAlaLeuLeuAspThrAlaAsnValArgHisIle 2124  
 Db 6730 AATGAGATATGCTGTGTGGCGGCTGTGCACTGTGCTCCCAAGACAGGAGGACTTGTG 6789  
 QY 2125 GlnLeuLeuGlnGlnThrGlnGly-----GlnThrAlaIlePheLeuGlnHis 2140

[illegible]

Qy	2495	GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSer	2514
Db	7930	GGCTTGTCTGTGGCCCTGGACCCCTGAAGGCGTAAAGGAACCTTAATTCTGTGGATCTCA	7989
Qy	2515	ILeTyrAspThrLeuLeuIleTyrPserPheAlaGlyProValAlaPheAlaValSerMetSer	2534
Db	7990	GTCCACAGGCCCTTCATCTGGAGCTTGTGGCCCTGTGTCTCGATCATGTATGATGAAC	8049
Qy	2535	ValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla---AlaGlnArgGlnGlyPhe	2553
Db	8050	GGGACCATGTTCTCTCTGGCTGGCCGACATCTGTCTCCACAGGGCAGAGGGAGGCC---	8106
Qy	2554	GlnIlybIysGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSer	2573
Db	8107	---AAGAAAGACCTTCGACATCGAACCTTCGGAGCTTCCTTCGTCGCTTCGTGGTGGTCA	8163
Qy	2574	AlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeu	2593
Db	8164	GGCTCCGGGCTCTTGTGGCTCTCCGGACAGTCAACACAGCATCTCAAGCTTCCTCACTTCC	8223
Qy	2594	PheAlaThrCysbAsnCybIleGlnGlyProPheIlePheLeuSerTyrValValLeuSer	2613
Db	8224	CATCTGGACTCTCGGCCCTCCAGAGGCTGGCGGTGGTGGTGGTCTCTTGTGTCTTAAT	8283
Qy	2614	LysGlnValArgIlybAlaLeuIlybLeuAlaCys---SerArgIlybProSerProAspPro	2632
Db	8284	GCAGATCTCGGGGCTGCTGGATGTCAGCTGTCTGTGGCAGAAAGAGCGCTCGAGAG	8343
Qy	2633	AlaLeuThrThrIlybSerThrIleuThrSerSerTyrbAsnCybProSerProTyrAlaAsp	2652
Db	8344	GCAAGGCCACAGCACTGGGCTGGGACCTGGGGCTTACAACAACACGGCTCTCTTGAAGAG	8403
Qy	2653	GlyArgLeuTyrGln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArg	2671
Db	8404	AGTGGCTCATTCGACATCATCTGGGGGCTCCACCGTCTCTGTGAAGAGAGCGCCGC	8463
Qy	2672	SerGlyIlybSerGlnPro-----SerTyrIle-----ProPhe	2688
Db	8464	TCCGGCCGAGCCACGAGACACAGACAGCCAGCGGGGCCGAGCTTCCACAGGACATATGC	8523
Qy	2683	LeuLeuArgGlnGluIleSerAla-----LeuAsnProGlyGlnGly	2695
Db	8524	CTGGTGTGCATGTGGCTCAGCGCGCTGACCACTACATCAACAGCTTCCAGGCTCATGTGGC	8583
Qy	2696	ProProGlyIleuGly-----AspProGlySerIleuPheLeuGlnGly	2709
Db	8584	CCCACTAGCTGACCTGGCAATGTCATCATGAGATGCTGGCGCA-----	8628
Qy	2710	GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAsp	2729
Db	8629	-----GACTCCGACTCTGACAGTACCTGTCTCTTGAAGAGAGAG	8667
Qy	2730	GlnSerGlySerTyrAlaSerThrHisSerSerAspSerArgGlnGlnGlnGlnGln	2745
Db	8668	AGAGTCTCTTCATCTTCATCTTCAGAAAGCAGAGACATAGCCGAGCGGGGGCGGCTC	8727
Qy	2750	GlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGlyProGlyAla	2765
Db	8728	CAAGGGCCACTCTGCCAGGAGCCAGCCAG-----AGT	8757
Qy	2770	GlnArgLeuProLeuHisSerThrProIlybAspGlyGlyProGlyIlybAlaAsp	2789
Db	8758	GAGAGGCTCTTCACCCAC-----CCCAAGATGTGAATGAGCAATGACCTCTGTCTAC	8811
Qy	2790	TrpProGly-----AspPheGlyThr-----	2796
Db	8812	TGGCCAGCGCTGGGGAGTGGAGGAGAGCCCTGTGCTGTGACAGACTTGGGGCTCTGAA	8877
Qy	2797	-----ThrAlaIlybGlnSerSerGlyAsnGlyAlaProGlnGlnArg	2810
Db	8872	AGGGCGCTGGGGCTGAGACCGAGAGAGATGACATTAACAACAACCAAGCCAGACCGCGGC	8931

```

Qy 2811 LeuArgGluuengLYAappAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySer 2830
Db 8932 CTG---ACCAgTGGGAT-----GAGACTTCTCTG-----GCCCG 8964
Qy 2831 SerAlaGlnProHisLysGlyIleLeuLysLysCys---LeuProThrIleSerGlu 2849
Db 8965 GCCCAGCGCCAGAGAAAGGCACTCTGAAGAACCGGTTGCCAATACCACTGGTCCACAG 9024
Qy 2850 LysSerSerLeuLeuArgLeuProLeuGluGlnCysThrGlySerSerArgGlySerSer 2869
Db 9025 ACCCGAGGTGCCCTCTGAG-----CTGTCTGTGTGCCGTGCAGCCACCTTGGGCCACCGT 9078
Qy 2870 Ala-----SerGluGlySerArgGlyGlyProProPro 2880
Db 9079 GCTGTGCCAGCTGCTCTTAAGGTGCACTTATGCTGCGGGGACACGGGACGCTTTCA 9138
Qy 2881 ArgProProProArgGlnSerLeuGlnGluLeuAsn 2893
Db 9139 CAGCCAGCCAGCCGCTACTCTTCTAAGAAACAGCTGAGC 9177

```

Search completed: February 13, 2004, 20:34:19  
 Job time : 2604 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 19:39:01 ; Search time 305 Seconds

(without alignments)  
4230.042 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545  
Sequence: 1 MRSPTAGVPLPTPPPLLL.....AGTVDBDSGSEPLFNNFLH 2923

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DRV=x1h  
-O=/cgn2\_1/USPTO.spool/US09916649/rnat/11022004\_143817\_20963/app.query.fasta\_1.3079  
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MIMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09916649 @CGN 1 1 169 @rnat 11022004\_143817\_20963 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEJOBBY -NEG SCORES=0 -WAIT -DSBELCK=100 -LONGLOG  
-DRV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database:

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1104	7.1	4566	2 US-08-465-976A-1	Sequence 1, Appl1
2	1104	7.1	4566	2 US-08-982-412-1	Sequence 1, Appl1
3	1032.5	6.6	4078	4 US-09-016-434-1066	Sequence 1066, Ap
4	881.5	5.7	4650	1 US-07-998-003A-102	Sequence 102, App
5	881.5	5.7	4650	1 US-08-453-274B-102	Sequence 102, App
6	881.5	5.7	4650	1 US-08-453-695A-102	Sequence 102, App
7	881.5	5.7	4650	1 US-08-268-161A-102	Sequence 102, App
8	881.5	5.7	4650	1 US-08-453-702A-102	Sequence 102, App
9	881.5	5.7	4650	2 US-09-039-639-102	Sequence 102, App
10	881.5	5.7	4650	3 PCT-US93-12588-102	Sequence 102, App
11	881.5	5.7	4650	5 PCT-US95-08071-102	Sequence 102, App
12	868.5	5.6	4104	1 US-07-998-003A-94	Sequence 94, Appl1

13	868.5	5.6	4104	1 US-08-453-274B-94	Sequence 94, Appl1
14	868.5	5.6	4104	1 US-08-453-695A-94	Sequence 94, Appl1
15	868.5	5.6	4104	1 US-08-268-161A-94	Sequence 94, Appl1
16	868.5	5.6	4104	2 US-08-453-702A-94	Sequence 94, Appl1
17	868.5	5.6	4104	3 US-09-039-639-94	Sequence 94, Appl1
18	868.5	5.6	4104	3 PCT-US93-12588-94	Sequence 94, Appl1
19	868.5	5.6	4104	5 PCT-US95-08071-94	Sequence 94, Appl1
20	780.5	5.0	5610	4 US-09-262-537-57	Sequence 57, Appl1
21	768.5	4.9	5693	4 US-09-262-537-19	Sequence 19, Appl1
22	733.5	4.7	5491	4 US-09-262-537-5	Sequence 5, Appl1
23	722	4.6	4190	4 US-09-262-537-9	Sequence 9, Appl1
24	722	4.6	5597	4 US-09-262-537-1	Sequence 1, Appl1
25	722	4.6	5598	4 US-09-262-537-33	Sequence 33, Appl1
26	697	4.5	2063	4 US-09-620-312D-1	Sequence 1, Appl1
27	691.5	4.4	5617	4 US-09-262-537-3	Sequence 3, Appl1
28	674	4.3	2072	4 US-09-620-312D-79	Sequence 79, Appl1
29	628.5	4.0	3033	1 US-08-453-695A-111	Sequence 111, App
30	628.5	4.0	3033	1 US-08-268-161A-111	Sequence 111, App
31	628.5	4.0	3033	2 US-08-453-702A-111	Sequence 111, App
32	628.5	4.0	3033	3 US-09-039-639-111	Sequence 111, App
33	628.5	4.0	3033	5 PCT-US95-08071-111	Sequence 111, App
34	627.5	4.0	4705	1 US-07-998-003A-96	Sequence 96, Appl1
35	627.5	4.0	4705	1 US-08-453-274B-96	Sequence 96, Appl1
36	627.5	4.0	4705	1 US-08-453-695A-96	Sequence 96, Appl1
37	627.5	4.0	4705	1 US-08-268-161A-96	Sequence 96, Appl1
38	627.5	4.0	4705	2 US-08-453-702A-96	Sequence 96, Appl1
39	627.5	4.0	4705	3 US-09-039-639-96	Sequence 96, Appl1
40	627.5	4.0	4705	5 PCT-US93-12588-96	Sequence 96, Appl1
41	627.5	4.0	4705	1 US-08-453-695A-109	Sequence 109, App
42	624.5	4.0	3353	1 US-08-453-702A-109	Sequence 109, App
43	624.5	4.0	3353	2 US-08-453-702A-109	Sequence 109, App
44	624.5	4.0	3353	2 US-09-039-639-109	Sequence 109, App
45	624.5	4.0	3353	3 US-09-039-639-109	Sequence 109, App

## ALIGNMENTS

RESULT 1  
US-08-465-976A-1  
; Sequence 1, Application US/08465976A  
; Patent No. 5869632  
; GENERAL INFORMATION:  
; APPLICANT: SOPPET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: RUBEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARIELLA, BYRNE, BAIN GILFILLAN, CECCHI  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY F  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-444  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 994-1744  
; TELEFAX: (201) 994-1744  
; INFORMATION FOR SEQ ID NO: 1:

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 4566 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE
: NAME/KEY: CDS
: LOCATION: 212..2863
: US-08-465-976A-1

```

Alignment Scores:	
Pred. No.:	2.38e-60
Score:	1104.00
Percent Similarity:	52.40%
Best Local Similarity:	36.93%
Query Match:	7.10%
DB:	2
Length:	4566
Matches:	277
Conservative:	116
Mismatches:	245
Indels:	112
Gaps:	22

US-09-916-849A-3 (1-2923) X US-08-465-976A-1 (1-4566)

QY	2203	lIleuPProgluSeValPheArglu-----ThrProValValArgProAla	2219
Db	26	GTCTGGCCCAAGAGCAGCATAGAAATCTCCACCACTCAAGTGTGGTCCCCCA	85
QY	2220	GlYProgluValaglIngluProgluIuIeuAlaArgArgInlArgArgHieProglu	2239
Db	86	GCCTCG-----CGAGG	97
QY	2240	LeuSerIngluIgluAlaValAlaSerValIleIleIYrArgThrLeuAlaIglYleuLeu	2259
Db	98	CCAGAGCCTGGAGATCCTCATATATATCTCCTCGTTACGGCACTTAAGGGGACTCTC	157
QY	2260	ProHIsenTyYrAbpProAbpLysArgSerIleuArgValProLysArgProIleIleIsn	2279
Db	158	CCTGCGCCAGTTCACGGCAGAAACGCCAGAGTCCACGGCTTCTCCAGAAACCCCGTATGAC	217
QY	2280	ThrProValValSerIleSerValHisAbpArgIngluIleuLeuProArgAlaLeuAsp	2299
Db	218	TCCCCGGGTGGTCAAGTGGCTGTGTTCCACGAGAGCACTTCTTAAGGGGAATCCTGGAG	277
QY	2300	LysProValThrValGlnPheArgPheuIngluThrIngluArgThrLysProIleCys	2319
Db	278	TCCCCCAATCAAGCTTAGAGTTTCGCTGCTTACAGACAGGAATCCGAGAGAGCATCTGT	337
QY	2320	ValPheTrpAsnHisSerIleLeuValSerGlyThrIglYlYrPserAlaArgGlyCys	2339
Db	338	GTGCAGTGGAGCCCACTGGCTGGCGGAGAGCATGTGTGTGGACAGCAGCGGACTGC	397
QY	2340	GluValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPhe	2359
Db	398	GAGCTGGTGCACAGAAATGGGTCCACCAACGATGTCTCTCAGCGGAGACGACTTT	457
QY	2360	AlaValLeuMetAspValSerArgArgGlu-----AsnGlyGluIleLeuProLeuLys	2377
Db	458	GGGGCTCTCAAGGAGGCTCTCCCTGAGAGGCGTGAAGGGCGACACTGAGACTGTGGCT	517
QY	2378	ThrLeuThrTyYrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhe	2397
Db	518	GTGTTCACCCACGTGTGGCTGTGTCTGTGGTGGCTGTGGCTGATCTGACGACCATC	577
QY	2398	LeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAla	2417
Db	578	CTGCTGAGCCTGGCGAGCCTCAAGTCCCAATGTGGTGGAAATCCATGCCAAATGTGGACCC	637
QY	2418	AlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPhe	2437
Db	638	GCCCTGGGGGGTGGCAGACTCTCTCTCTGCTGGGGAATTCACAGAACCCAAATCAAGCTG	697
QY	2438	AlaCysThrValIleAlaIleLeuLeuHisPheLeuTyYrLeuCysThrPheSerTrpAla	2457
Db	698	GTGTGCATCGCAGTGGCCATCTCTCTGACATCTTCTCTCAGACCTTGCGCTGGTCTC	757

QY	2458	LeuLeuGluAlaLeuHisLeuYrYrArgAlaLeuThrGluValAlaArgAspValaAsnThrGly	2477
	.....		
Db	758	TTCTGTCAGAGGGCGCTGCACCTTCAACCGATGACAGGTTGAGCCACCGCAACGTGGACCGCGGC	817
QY	2478	ProMetArgPheYrYrYrMetLeuGlyYrTrpGlyValProAlaPheIleThrGlyLeuAla	2497
	.....		
Db	818	GCCATGGCGCTTCAACATGCGCTTGCGGTGGGGGGCTCCCTGCTGTGCGTGGCGCTTGCTGCT	877
QY	2498	ValGlyLeuAspProGluGlyYrGlyYAsnProAspPheCysTrpLeuSerIleYrYrAsp	2517
	.....		
Db	878	GTGGGCGCTGCAGCCTTGAGGGCGCTATGGGAGACCTGCATCTTGTGGTAATCACTCACTCAAGAG	937
QY	2518	ThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeu	2537
	.....		
Db	938	CCCTCATCTTGAGACTTGTCTGCGCCCTGTGTTCTGCTGTCTCATGATGAAACGGACCATG	997
QY	2538	TyrIleLeuAlaAlaArgAlaSerCysAla--AlaGlnArgGlnGlyPheGluYrAs	2556
	.....		
Db	998	TTTCTCTCTCGCGTCCCGCCACATCTGTCTTCCACAGGGCAGAGGGAGGCGC-----AAGAG	1051
QY	2557	GlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrp	2576
	.....		
Db	1052	ACCTTCGACTGACCTTCGGACCTCTTCTGTGTCTTGTCTGTGCTGAGTACAGTCTCTGG	1111
QY	2577	LeuLeuAlaLeuLeuSerValaAsnSerAspThrLeuLeuPheHisYrLeuPheAlaThr	2596
	.....		
Db	1112	CTCTTTGGGCGTCCCTGGAGCTCAACACAGATCTTACCTTCACTCACTCATGCTGGA	1171
QY	2597	CysAsnCysIleGlnGlyProPheIlePheLeuSerYrValValLeuSerYrGluVal	2616
	.....		
Db	1172	CTTCGCGCGCTCCCGAGGCGCTGGCGGCTGCTGCTCTGTGTGCTTAAATGACAGATCT	1231
QY	2617	ArgYrValaLeuYrSleuAlaCys---SerArgYrProSerProAspProAlaLeuThr	2635
	.....		
Db	1232	CGGGCTGTGATGCCAGCGCTGTCTGGGCGAGAGGACAGCGCTTGAGAGGACAGGCCA	1291
QY	2636	ThrYrSerThrLeuThrTrpSerYrYrAsnCysProSerProYrYrAlaAspGlyArgLeu	2655
	.....		
Db	1292	GCACCTGAGGCTGGACCTGGGCGCTTCAACAACAGCGCTCTTTGAGAGAGAGTGGCTTC	1351
QY	2656	TyrTrpIn---ProYrGlyYrAspSerAlaGlySerLeuHisSerThrSerArgSerGlyYr	2674
	.....		
Db	1352	ATCCGCACTCACTCGGGCGCTCCACCGCTCTCTGTGAGCAAGTGGCCGGCTCGGGCGG	1411
QY	2675	SerGlnPro-----SerYrIle-----ProPheLeuLeuArg	2685
	.....		
Db	1412	ACCGAGACCAAGACAGACAGCGGGCGCGGACGTACTCAAGGGACATATGCTCTGTGGA	1471
QY	2686	GluGluSerAla-----LeuAsnProGlyGlnGlyProProGly	2698
	.....		
Db	1472	CATGGCTCAAGCCGCTGACCACTGACCAAGCGCTCAAGCGTCAATGCTGAGCCCACTGAC	1531
QY	2699	LeuGly-----AspProGlySerLeuPheLeuGlnGlyGlnaAspGln	2712
	.....		
Db	1532	CTGACAGCTGACATGTTCCATCGAATGCTGGGCA-----	1567
QY	2713	GlnHisAspProAspTrpAspSerAspSerAspLeuSerLeuGlnaAspAspGlnSerGly	2732
	.....		
Db	1568	-----GACTCCGACTCTGACAGTGAACCTGTCTTGGAGGAGAGAGGAGTCTC	1615
QY	2733	SerYrYrAlaSerThrHisSerSerAspSerGluGluGluGluGluGluGluGlu	2752
	.....		
Db	1616	TCATATTCATTTTGAAAGAGAGACATATGGCCGACCGGGGGCGCTTCCACAGGCCA	1675
QY	2753	AlaAlaPheProGlyGlnGlnGlyYrTrpAspSerLeuLeuGlyProGlyAlaGlnaArgLeu	2772
	.....		
Db	1676	CTTCGCGAGCGAGCGAG-----AGTGAAGAGCTCTC	1705
QY	2773	ProLeuHisSerThrTrpProYrAspGlyGlyProGlyYrYrAlaProTrpProGly	2792
	.....		
Db	1706	CTCACCCAC-----CCCAAGAATGTGAATGGCATGACCTCTGTCTTACTGCGCAGCC	1759
QY	2793	-----AspPheGlnYrTrp-----	2796

```

Db 1760 CTGGGGAGTGGCGAGCGCCCTGTGCTCTGCAAGCTTGGGGCTCTGAAAGCGCTTG 1819
Qy 2797 -----ThralatylglsuserSerglyAsnlyalProglugluArgleuArglu 2813
Db 1820 GGGGTGACACACGACAGATGACGCTACACACACAGCCAGCCCGGCTTG--ACC 1876
Qy 2814 AsnlyalAspAlaLeuSerArggluglySergleuylProleupProgllySergSerAla 2833
Db 1877 AGTGGGGAT-----GAGACTTCTCTG-----GGCCGGGCCCAAGCC 1912
Qy 2834 ProHialysgllyleuyllylylCys---LeuProthrllesergluylsSerSer 2852
Db 1913 CAGAGAAAGGACCTCTGAAAGAACGGTTCATATACCTACTGTGACACAGCCCGAGGT 1972
Qy 2853 LeuLeuArgleupProleugluCysThrlySergSerArglySergSerAla----- 2870
Db 1973 GCCCGTAG-----CTGCTCGGTGCGCGAGCCACCTTGGGCCACCGTCTGTGCCA 2026
Qy 2871 -----SergluylSergArglylylProProProArgProPro 2883
Db 2027 GCTGCTCTTACCGTCCATCTATGCTGGCGGGGACAGGGGACCTTTCAAGCCAGCC 2086
Qy 2884 ProArgInSerleuIngluInleuAsn 2893
Db 2087 AGCCGCTACTCTTCTAGAGAACAGCTGGAC 2116

RESULT 2
US-08-982-412-1
; Sequence 1, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE,
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,412
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF181PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 212..2863
; US-08-982-412-1

```

```

Alignment Scores:
Pred. No.: 2,38e-60 Length: 4566
Score: 1104.00 Matches: 277
Percent Similarity: 52.40% Conservative: 116
Best Local Similarity: 36.93% Mismatches: 245
Query Match: 7.10% Indels: 112
DB: 2 Gaps: 22

US-09-916-849A-3 (1-2923) x US-08-982-412-1 (1-4566)
Qy 2203 lIeLeuProgluSerValPheArglu-----ThProProValValArgProAla 2219
Db 26 GTCTCCCAACAGACAGCAGCATAGAAATCCACCACTCAAGTGTGTCCTCCCAACA 85
Qy 2220 GlYProglYgluAlaIngluProgluInleuAlaArgArggluArgArgHISProglu 2239
Db 86 GCCCGG-----CCAGAG 97
Qy 2240 LeuSerInglYgluAlaValAlaSerValIleIleTYArgThrleuAlaGlyleuLeu 2259
Db 98 CCAAGCCCTGGAGATCCCATTTAATTCCTCTGTTACCGACCTTAGGGGAGCTGCTC 157
Qy 2260 ProHialAsnTYAspProAspIlyAspSerleuArgValProlyAspProIleleuAsn 2279
Db 158 CTTGCCAGTTCCAGGACAGACGCCAGGTGCGAGGCTTCTCAAGAACCCGTCATGAAC 217
Qy 2280 ThrProValIaSerIleSerValInAspAspIngluInleuProArgAlaLeuAsp 2299
Db 218 TCCCGGTGTGTAGGTGTGTGTGTCCAGGACCACTTCTTAAGGGGAATCTGTGAG 277
Qy 2300 LySProValIthValGlnPheArgleuIngluIngluArgThrlyAspProIleCys 2319
Db 278 TCCCCATCAGCCTAGATTGCGCTGACAGACAGCAATCGAGGAGCAAGCGATCTGT 337
Qy 2320 ValPheThrAsnHISerIleleuValSerglYthnglylylTYPSerAlaArgglYcys 2339
Db 338 GTGCAGTGGAGCCACCTGCGCTGGGAGACAGCATGTGTGTGTGACACAGGAGCTGC 397
Qy 2340 GluValValPheArgAsnInuSerHISValSerCysGlnCysAsnHISerThSerPhe 2359
Db 398 GAGTGTGTCACAGGATGGTCCACGCGAGGTGTGCTGAGCGGACAGGACCTTT 457
Qy 2360 AlaValleuMetAspValSerArgglu-----AsnlygluIleuProleuIys 2377
Db 458 GGGTCTCATATGATGCTCTCCCGTAGAGCTGAGGGGACCTGAGCTGGCTGGCT 517
Qy 2378 ThrleuThrValAlaLeuGlyValThrleuAlaAlaLeuLeuThrPhePhe 2397
Db 518 GTTTCACCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 577
Qy 2398 LeuThrleuLeuArgIleleuArgSerAsnGlnHISglYIleArgArgAsnleuThrAla 2417
Db 578 CTGTGAGGCTGCGAGGCTGACATGCAATGCGGTGGATCCAGGCCAATGGGAGGC 637
Qy 2418 AlaLeuGlyleuAlaGlnleuValPheleuGlyIleAsnGlnAlaAspLeuProPhe 2437
Db 638 GCCCTGGGGGTGACAGGTCTCTCTGCTGGGAGATTCACAGACCCACATGACGTG 697
Qy 2438 AlaCysThrValIleAlaIleleuLeuHISpHeleuTYrleuCysThrPheSerTrpAla 2457
Db 698 GTGTGACCTGACGTGCGCATCTCTGACATCTTCTTCTGACACCTTCCGCGGTGC 757
Qy 2458 LeuLeuGluAlaLeuHISleuTYrArgAlaLeuThrGluValArgAspValAsnThrIy 2477
Db 758 TTGCTGAGAGGGGTGACCTTACCGCATGACGTTGAGCCACGACAGCGGAGCG 817
Qy 2478 PrometArgPheTYrTYrMetleuGlyTYrGlyValProAlaPheIleHISlyleuAla 2497
Db 818 GCCATGCGCTTCTACCATGCGCTGGGCGGTGCTGCTGTGTGTGTGTGTGTGTGTGT 877
Qy 2498 ValGlyLeuAspProgluGlyTYrGlyAsnProAspPheCysTYrPleuSerIleTYrAsp 2517
Db 878 GTGGGCTGTGAGCCCTGAGGGGTATGGGAACCTGACTTGTGTGATCTCACTCAAG 937

```

```

QY 2518 ThrLeuIleTPSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeu 2537
Db CCCCTCATCTGAGAGCTTGTGTGGCCCTGTGTGCTGTGATAGTATGAGGAGCCATG 997
QY 2538 TyrIleLeuAlaAlaArgAlaSerCysAla---AlaGlnArgGlnGlyPheGlnLys 2556
Db TTTCTCTCTGCTGCGCCGACATCTGCTGCACAGGAGGAGGAGGAGCC---AAGAG 1051
QY 2557 GlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyr 2576
Db ACCCTGTGACAGACCTTCGAGCTCTTCTGCTGCTTCTGCTGCTGCTGCTGCTG 1111
QY 2577 LeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheLeuSerTyrLeuPheAlaThr 2596
Db CTCTTGGGCTCTCTGAGATCAACACAGCATCTTCACTTCACTTCACTTCACTTCA 1171
QY 2597 CysAsnGlyIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlnVal 2616
Db CTCTGCGGCTCTCAAGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1231
QY 2617 ArgLysAlaLeuLysLeuAlaCys---SerArgLysProSerProAspProAlaLeuThr 2635
Db CGGCTGCTGCTGATGCCAGCTGTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1291
QY 2636 ThrLysSerThrLeuThrSerSerTyrAsnGlyProSerProTyrAlaAspGlyArgLeu 2655
Db GCACCTGGGCTGGGAGCTGGGAGCTTCAACAACAGCGCTCTTGTGAGAGAGTGGCTC 1351
QY 2656 TyrGln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerLys 2674
Db ATCCGATCACTCTGCGGCTGCTGCACCGCTCTCTGAGAGAGTGGCTGCGGCGG 1411
QY 2675 SerGlnPro---SerTyrIle---ProPheLeuLeuArg 2685
Db ACCGAGACAGACAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1471
QY 2686 GlnGlnSerAla---LeuAsnProGlyGlnGlyProProGly 2698
Db CATGCTCAGCGCTGACACACATGACACAGCTTCCAGGCTCATGCGCCCACTGAC 1531
QY 2699 LeuGly---AspProGlySerLeuPheLeuGlnGlyGlnAspGln 2712
Db CTGACGCTGACATTTCCATGAGATGCTGCGCA--- 1567
QY 2713 GlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAspGlnSerGly 2732
Db GATCTCGACCTTCAAGATGACCTGTCTTGTGAGAGAGAGAGAGAGTCTC 1615
QY 2733 SerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnGlnGlnGlnGln 2752
Db TCCATTTCACTTCAAGAAAGCAGACATGAGCCGAGCGGCGGCTTCCAAAGGCCA 1675
QY 2753 AlaAlaPheProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2772
Db CTCTGCCAGACAGGCGCAG---AGTAGAGGCTC 1705
QY 2773 ProLeuHisSerThrProLysAspGlyGlyProGlyProGlyLysAlaProTyrProGly 2792
Db CTCACCCAC---CCCAAAGATGATGATGATGATGATGATGATGATGATGATGATG 1759
QY 2793 ---AspPheGlyThr--- 2796
Db CTGAGGAGATGAGAGAGCCCTGTGTCTGTGACAGATGAGGCTGTGAAGAGCGCTG 1819
QY 2797 ---ThrAlaGlyGlnSerSerGlyAsnGlyAlaProGlnGlnArgLeuArgGln 2813
Db GGGCTGACACAGACAGAGATGACATCAACACAGACAGACCGGCTG---ACC 1876
QY 2814 AsnGlyAspAlaLeuSerArgGlnGlySerLeuGlnGlyProLeuProGlySerSerAlaGln 2833
Db AGTGGGAGT---GAGACTTCTCTG---GGCGGAGCCAGGCGC 1912

```

```

QY 2834 ProHisGlyGlyIleLeuLysLysCys---LeuProThrIleSerGlySerSer 2852
Db CAGAGAAAGGACATCTCGAAGAACCGGTTCAATACCACTGGTGGCAGACCCGAGT 1972
QY 2853 LeuLeuArgLeuProLeuGlnGlnGlyThrGlySerSerArgGlySerSerAla----- 2870
Db GCCCTGAG-----CTGTCTGTGTGCGTGGACCACTTGGGCGCACCGTGTGCGCA 2026
QY 2871 ---SerGlnLysSerArgGlyLysProProArgProPro 2883
Db GCTGCCCTTAAGGTGACATTAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2086
QY 2884 ProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2893
Db AGCGCTACTCTTCTAGAGAAACAGCTGAC 2116

RESULT 3
US-09-016-434-1066
Sequence 1066, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
Applicant: Janice Au-Young
Applicant: Jeffrey J. Sellhammer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESSES:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1066:
SEQUENCE CHARACTERISTICS:
LENGTH: 4078 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g1107686
US-09-016-434-1066

Alignment Scores:
Pred. No.: 6,96e-56 Length: 4078
Score: 1032.50 Matches: 280
Percent Similarity: 47.85% Conservative: 132
Best Local Similarity: 32.52% Mismatches: 348
Query Match: 6.64% Indels: 101
DB: 4 Gaps: 23

US-09-916-849a-3 (1-2923) x US-09-016-434-1066 (1-4078)

```

QY 179 ProGlnPheGlnProSerTyrGlnAlaThrValProGluGlnProAlaGlyThr 198  
DB 1567 CCTGAATTTCCAGACAGACGGTCAAGCTGTTTGCATGAGAACGCGCATTTGGACT 1626  
QY 199 ProValAlaSerLeuArgAlaIleAspProAspGlnGlyValAlaGlyLeuGlyTyr 218  
DB 1627 ACTATGATGAGCTGAGCTGATGACCTGATGAGGAGTGAAGATGGTATCGTGAACATAC 1686  
QY 219 ThrMetAspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThr 238  
DB 1687 AGATATGCAAAATTTA-----AATCATGTGCGCTTGTGCGATGACATTTCACT 1734  
QY 239 GlyAlaValThrAlaGlnGluLeuAspArgGlu---ThrLysSerThrIleValPhe 257  
DB 1735 GGTGCGGTGAGTACGTACAGAAACCTGCACTACGAACCTGATGCGGTATTATACCTG 1794  
QY 258 ArgValThrAlaGlnAspHisGlyMetPro---ArgArgSerAlaLeuAlaThrLeuThr 276  
DB 1795 AGGATTCGTGCATCAGACTGGGGCTTGGCCGTACCGCGGGAAGTCCGAAGCTCTTGCTACA 1854  
QY 277 IleLeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnGlyTyrLysGlu 296  
DB 1855 ATTACTCTCAATTAATGGAATGACACACACTTTGTTGAGAAATTAATTTGTAAGG 1914  
QY 297 SerLeuArgGlnLeuLeuGlnValGlyTyrGluValLeuThrValAlaGlnAlaThrAspGly 316  
DB 1915 ACAATTCGCCAGATCTAGCGGTGGAGCAATTAACCACTCTTCTGCTATGTATGCA 1974  
QY 317 AspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyLysSerPro 336  
DB 1975 GAT-----GACTTCAGTTGTGTACAGTACATCAAGATTGA-----GCTGGAATAAGAA 2019  
QY 337 SerGlnValPheGlnIleAspProArgSerGlyValIleArg---ThrArgGlyProVal 355  
DB 2020 CTGATCTTGTGTTGTTAAACCCCACTCGGGGTATGTGATTAAGAAGTCGCTAAG 2079  
QY 356 Asp-----ArgGlnGlnValGlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGly 373  
DB 2080 GATGGCTTAGGTGCAAGGTGTCTTTCACAGCTGATGATCAACACTACATCAAGATGAGAA 2139  
QY 374 ArgAspProGlyPro-----ArgSerThrAlaIle----- 384  
DB 2140 AATTTCGCCACACCATTAATATCAACATTAACAGTGCCTCCAGTCAACAGCTGTAAAC 2199  
QY 384 ----- 384  
DB 2200 TTGCAGTGTGAGAGACTGGTGTGGCAAAATGCTGGCAGAGAAAGCTCCGACGCAAAAT 2259  
QY 385 -----ValPheLeuSerValGlnLeuAspAsnAsp 394  
DB 2260 AAATTAACAACAACGAGGAGAGTGAAGATATTTCTTCGATTCTCACTCTGCAATGCT 2319  
QY 395 AsnAlaProGlnPhe---SerGlnLysArgTyrValIleGlnValArgIleAspValThr 413  
DB 2320 CACATACCGCAGTTTAGAGCACTCTCCAGCTGATTAATGAGTAAAGGAAACCAAGCT 2379  
QY 414 ProGlnAlaProValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaVal 433  
DB 2380 GTGGGTTCACGATGATTTTCACTGACCTCACTGACCTGACCTGCTTCATATGAAAA 2439  
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
DB 2440 CTGTGCTATGCTTCTTTCGAGGAATGAGATAGTTCATCATGATGATAGAAAAA 2499  
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGlnThrLysGlnTyrThrLeuArg 473  
DB 2500 GGAATGCTGAATAATTTATCTCTCTTGAACCGTGAACCAACACAAATTAACCTGAAAT 2559  
QY 474 ValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
DB 2560 ATTAACGCTATGACCTTGGGATACCCCAAGAGCT---GCGTGGCGTCTTCTCAATGTC 2616

QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
DB 2617 GTGGTGTGATGATCCCAATGATATATCAACCGAGTTTTCACAGAGCATATTTTGGA 2676  
QY 514 ValLeuGlnSerValProLeuGlyTyrLeuValIleHisValGlnAlaIleAspAlaAsp 533  
DB 2677 GTGAGTGAAGACAGAGGATGATATGTAATCAATCCAGGTGAAGCCACAGATTAAGAC 2736  
QY 534 AlaGlyAspAsnAlaArgLeuGlnTyrArgLeuAlaGlyValIleGlyHisAspPheProPhe 553  
DB 2737 CTGGGGCCCAAGGACAGCTACGTACATTAAT-----CTTACAGACACACACATTT 2790  
QY 554 ThrIleAsnAsnGlyThrGlyTyrPheSerValAlaAlaGlnLeuAspArgGlnGlnVal 573  
DB 2791 TCAATTAACAGCGTACCGGCTGTGTTAAATCCAGACCGCTCTGATGAGAGCTGACG 2850  
QY 574 AspPheTyrSerPheGlyValGlnAlaArgAspHisGly-----ThrProAlaLeuThr 591  
DB 2851 CATGACACTCTTAAGATGAGCCAGGAGCCAGACCAAGCAAGAAAGAGCTGAGCTGTC 2910  
QY 592 AlaSerAlaSerValSerValThrValIleAspValAsnAspAsnProThrPheThr 611  
DB 2911 TCCACTGTGCTGTGAAGATACACTAGAAAGATGTTATGACACCACTCACTTAT 2970  
QY 612 GlnProGlyTyrThrValArgLeuGlnLeuAspAlaAlaValGlyThrSerValValThr 631  
DB 2971 CCACCTTAATTAATCGTGTGAAGCTCCAGAGATCTTCCAGAAAGAACCTCATCATGTGG 3030  
QY 632 ValSerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsn 651  
DB 3031 TTAGAAAGCCACGATCCGTGATTTAGTCA-----TGTGTGACG 3069  
QY 652 ThrArgAsnArgPheSerIleThrSerGlnSerGly----- 664  
DB 3070 GAG-----AGATACAGCTTGTGACCAAGGAAAGAACTTGAGTGTATTAATC 3123  
QY 665 ---GlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGlnArgGlnTyrValLeu 683  
DB 3124 AGTGAAGCAGTTAGGATGTCCAGCAGTGTGACCTTGAAGAACCAAGTGTATATCTC 3183  
QY 684 AlaValThrAlaSerAspGlyThrArg-----GlnAspThrAlaGlnIleVal 699  
DB 3184 ACTGTGAGGCGCAAGACAGGAAAGCAAGTTCCTGTCTTCACTTGTGATGTGAA 3243  
QY 700 ValAsnValThrAspAla-----AsnThrHisArgProValPheGlnSerSerHisTyr 717  
DB 3244 GTTGAAGTGTGATGTGAATGAGAACTGCACCAACCGTGTTCAGCTTGTGGA 3303  
QY 718 ThrValAsnValAsnGlnAspArgProAlaGlyThrThrValValLeuIleSerAlaThr 737  
DB 3304 AAGGGAACAGTGAAGAAAGATGCACCTGTGGTTCATTTGATGACGGTGTGCTCAT 3363  
QY 738 AspGlnAspThrGlyGlnAsnAlaArgIleThrTyrPheMetGluAsp-----SerIle 755  
DB 3364 GATGAGAGCGCGGAAGATGGGAGATCCGATCTCCTCATTTAGAGATGCTGTGGCTT 3423  
QY 756 ProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeuAspTyr 775  
DB 3424 GGTGTTTCAAAATAGTGAAGACAGGTGTCATAGAGACGTGCAGATCGACGTGACCGT 3483  
QY 776 GlnAspGlnValSerTyrThrLeuAlaIleThrAlaAspAsnGlyLysProGlnLys 795  
DB 3484 GATGACCTCCCATTAATTTGGCTAACATCTTGGACCGAAGAGGGGTCTGCTGCTT 3543  
QY 796 SerAspThrThrTyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPhe 815  
DB 3544 TCATGCTTCATAGATATCTACATGAGTTGAGATGTCATAGACAAATGCAACCAACAGCA 3603  
QY 816 LeuArgAspSerTyrGlnGlySerValTyrGlnAspValProProPheThrSerValLeu 835  
DB 3604 TCAGAGCTGTTTATTAACCAAGAAATCATGAAATTCCTTAAGATGATCTGTGGTC 3663  
QY 836 GlnIleSerAlaThrAspArgAspSerLysLeuGlnArgValPheTyrThrPheGln 855

[illegible]

i		LOCATION: 495..4103	
US-07-998-003A-102			
Alignment Scores:			
Pred. No.:	3,25e-46	Length:	4650
Score:	881.50	Matches:	343
Percent Similarity:	38.23%	Conservative:	146
Best Local Similarity:	26.82%	Mismatches:	452
Query Match:	5.67%	Indels:	339
DB:	1	Gaps:	46
US-09-916-849A-3 (1-2923) x US-07-998-003A-102 (1-4650)			
QY	3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeuLeu 22		
DB	513 AGCCCAAGCCCTGGGGGGGCAACGGCTACTGTGCTCCCTCCATGCTGCTTAGCATCTGCTC 572		
QY	23 LeuLeuLeuProProProProLeuLeuGlyAspIValGlyProCybArgSerLeuGlySer 42		
DB	573 CTGCTGGCTCATCCCAAG-----CCACGCCACTGGGTAGTGTGA 613		
QY	43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCybProSerSerAla 62		
DB	614 CAAGGTCCCGGAGGAMACGCCACCAACCTC---CATTTG-----GACCTCGC 661		
QY	63 SerAsnLeuTrpLeuTyrrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82		
DB	662 AGCCGACTTAGGTTT-----TCCAGATGTGGGGCACCTGTACAAAGCTAGAGGTGGTGC 715		
QY	83 LeuValPro-----HisHisAspGlyLeuArgVal 92		
DB	716 CCGCTACCTTCGGGTGATGGCAAGACAGGTGACATTTTCAACCAACCGAGACCTC----- 769		
QY	93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110		
DB	770 -----CATCGACCGGAGGGGCTCCGTGATGCGGAGAACCAAGCT 808		
QY	111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGluGlyLysLeuThr 130		
DB	809 CCCCTGGTGAATCCCTGCATCCCTGGAGTTTAGGT----- 841		
QY	131 LeuProGluGluHisProCysLeuLysAlaProArgLeuArgCysGlnSerCysLysLeu 150		
DB	842 -----ATCTATCAACAGACTGTGCAGATGC----- 868		
QY	151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyGlyArg 170		
DB	869 ---GAGCCCCCGGTCTGAAGGGCCAGATGAAGT----- 901		
QY	171 ArgLysArg-AsnValAsnThrAlaProGlnPheGlnProProSerTyrrGlnAlaThrVal 190		
DB	902 ACAAGACATCATATACAAACACA-----CCCACTTGCCCTCAACATCATCATCTGGGCAT 958		
QY	190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206		
DB	959 CCGTGAAGAACCAACATCGGCTCATCTTCCCATCCCGGTGGCTTCA----- 1007		
QY	206 eAspProAspGluGlyGluAlaGlyArgLeuGlyTyrrThrMetAspAlaLeuPheAsp 226		
DB	1008 -GACCGTGAATGCTGTGCCAACGGGTGGGACTCTTAAGTGCAGGTGGCGAGAGAACCA 1066		
QY	226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246		
DB	1067 GGAGGAGAGACA-----CCACAGCTCATTTGATGGGGCA 1102		
QY	246 uLeuAspArgGlyIleThrLysSerThrHisValPheArgValThrAlaGlnAspHisGly 266		
DB	1103 CCTGAGACCGTGAAGCGCTGGGACTCTTATGACCTCAACATCAAGGTGCGAGATGGCGGAGC 1162		
QY	266 tProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAspHisAs 286		
DB	1163 CCCCCACGGCCACGAGTGGCCCTGCTGCGGTGTACACGGTGTGTACCAATATGACACGGC 1222		





```

Db 2297 -----GCCCTGGCGTTCTTGTCGCTACTGCAGACAGGCGGAGCCAAAGCTGTGA 3049
Qy 983 -----AspleuAspTyrGluAspArgProGluTyrValLe 994
Db 3050 CCAGGCTGTGAAGAGAGACCAAGACCTG---TATCCCCCAAGCCCACTGCAAGGC 3106
Qy 994 uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014
Db 3107 CTCCAAGGAAACAACAAGCAAGAGCAAGAGCAAGTCCCAAGCCCGTGAAGCAGT 3166
Qy 1014 uAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAlaTyrVa 1034
Db 3167 GAGAGACAGAGATGAGCGCGGAGTCCAGAGTCCCTCAAGTTC----- 3209
Qy 1034 LThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054
Db 3210 ---AACTGATGAGCAGTACCGCTGGGAGACAGT---CCCCGACATC-----CACCTGCC 3256
Qy 1054 oAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLe 1074
Db 3257 C----- 3257
Qy 1074 uLeuAsnAlaSerThrArgIleuLeuValLeuSerArgAlaLeuAsnAsnArgProLe 1094
Db 3258 -CTCACTACCAACCAAGGAGCCCTGACCTGGCGCGCACTATGCTTACTCCCACT 3316
Qy 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111
Db 3317 GCCTTCATTCAGCTGACGACCCCACTCACTCAAGCTCCCAAGAGCAAGCAGGAGTGA 3376
Qy 1111 aGln-CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131
Db 3377 GAGCTGCGACCTGCA-----ACACATTCG----- 3402
Qy 1131 euArgLeuGlnAspMetSerPro-----GluArgPheLeuSerProLeuLeuG 1147
Db 3403 -----TGGGACCGGGGAGACCAAGTCCAGCGGCTGTGAGCAGT----- 3441
Qy 1147 ILeuPheIleGlnAlaValAlaIleThrLeuAlaThrProProAspHisValValAlP 1167
Db 3442 -----ACTCGAGTACAGCTACCGACCAAGCCCA----- 3474
Qy 1167 heAsnVal-GlnAspArgThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186
Db 3475 --AAATACCCAGCAAGCAAGTATCACTACCGCGCGCTCACTCTC----- 3517
Qy 1187 ValGlnIleProGlyProGlyGlyGlyProProPheLeuProSerGlu 1203
Db 3518 ---GCCCAACGACGAGGCCCAAGAGCTGACAGACCC---ATCCACAGCAG 3562

```

```

ATTORNEY/AGENT INFORMATION:
NAME: No. 5663300and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32660
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TRIALX: 25-3856
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 4650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 495..4103
US-08-453-274B-102

Alignment Scores:
Pred. No.: 3,256-46 Length: 4650
Score: 881.50 Matches: 343
Percent Similarity: 38.23% Conservative: 146
Best Local Similarity: 26.82% Mismatches: 452
Query Match: 5.67% Indels: 339
DB: 1 Gaps: 46

US-09-916-849a-3 (1-2923) x US-08-453-274B-102 (1-4650)
Qy 3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeu 22
Db 513 AGCCAGGCGCTGGGGGAGCAAGCGCTACCTGCTCCATGCTGTAGCAGCTGCTC 572
Qy 23 LeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42
Db 573 CTGCTGCTCATCCCAAG-----CCAGCCACTGGGTAGTGA 613
Qy 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysProSerSerAla 62
Db 614 CAAGTCCGGAAGAAACAGCACCAACCTT---CATTTG-----GAGCTCGC 661
Qy 63 SerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82
Db 662 AGCCGATATGTTT-----TCCAGATGGGGCACTGTACAGCTAAGGTGGTGC 715
Qy 83 LeuValPro-----HisIleAspGlyLeuArgVal 92
Db 716 CCCGACTTCCGCTGAGAGGCAAGACAGTGCATTTCACACCAAGACCTC----- 769
Qy 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
Db 770 -----CATCGACGTTAGAGGGCTCCGTAATGCGAAGACAGCT 808
Qy 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleLeuSerProGluGlyGlyLeuThr 130
Db 809 CCCGTGATCCCTGCATCTCGAGATTGAGT----- 841
Qy 131 LeuProGluGluHisProCysLeuValAlaProGluLeuArgCysGlnSerCysValLeu 150
Db 842 -----ATCTATCACAGACCTCTGCAAAATTC----- 868
Qy 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyGlyArg 170
Db 869 ---GAGCCCGCGCTGTAGAGGGCCAGATAGAGT----- 901
Qy 171 ArgIysArg-AsnValAsnThrAlaProGluInpGlnProProSerTyrGlnAlaThrVa 190
Db 902 ACAAGACATCAATGACACACA-----CCCAACTGGCTCAACAGTATCACTGGGCAAT 958
Qy 190 IProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206
Db 959 CCTGAGAACACCAACAGTGGCTCACTTCCCATCCGCTGCTCA----- 1007

```

QY	206	eaapPProaaprglunlglylalaGlyAargleuglufuryThrmetAapralaleuPheaspse	226
Db	1008	-ACCCGTATGTGTGGTCCCAAGCGTGTGGCATCTTATGAGCTGGCAGGTGGCAGAGACCA	1066
QY	226	rArgserAnglnPhePheSerLeuaspProvalThrGlyAlaValThrThrAlagluGl	246
Db	1067	GGAGGAGGAAGCA-----CCACAGCTCATTTGATGGGCA	1102
QY	246	uLeuapArglunThrLysSerThrHisValPheargValThrAlaglnAphHisglYme	266
Db	1103	CTGGAGCCGTGAGCCCTGGGACCTCTATGACCTGCACATCAAGGTGAGGATGGCGGAG	1162
QY	266	rProArgnrserrAlaleuAlaThrLeuThrLleuValThrApThrAenAphHis	286
Db	1163	CCCCCAGCGCCGACAGAGTCCCTGCTGGTGTCAACGTCTTGCACCAATGACCAAGC	1222
QY	286	pProvalPheglunGlnGluufuryLysgluserLeuArgluAenLeuGlyAlaGlyTY	306
Db	1223	CCCCAAGTTTAGGGCCCTCTATGAGGCCAATATCTGAGATACCCCATTTGGCCA	1282
QY	306	rgluValLeuThrValAargAlaThrAapGlyAapAlaProProAenAlaenIlleuTY	326
Db	1283	CTGGCTGATCACTGAGGTGAAGGCCCAATGACTCAGACCAAGTGCCAATGCAAAATGAAATA	1342
QY	326	rArgLeuLeuGlnGlyserGlyLyserProserGluVal-----PhegluIlleas	343
Db	1343	CACATTTCCACCAAG-----GGCGCCCAAGTTGTAGGCGCTCTTCTTCACGTGA	1390
QY	343	pProArgserGlyValIlleargThrAargGlyProValAapArglunGluValGlyserTY	363
Db	1391	CAGGAACACTGGACTTATCACTGTTCAAGGGCCGGTGAACGTGAGACCTTAAGACCCCT	1450
QY	363	rgluLeuThrValGluAlaSerAapGlnGlyAArgAapProGlyProArgSerThrAla	383
Db	1451	GGGCTTCACAGCTCTTGCTTAAGACCCAGGACCAAC-----CCCAAGATGGCCCGTGC	1504
QY	383	aAlaValPheLeuSerValGluAapAapAapAapAapAlaProGlnPhe-----	399
Db	1505	CCAGGTGTGTGACCGTGAAGAGCATGAATGACATCCCCCAACATTGAGATCCGGGG	1564
QY	400	-----SerGluLysAArgTYrValValGlnValArgLysAapValThrProGl	415
Db	1565	CATAGGCGTAGTGCATCAACAGATGGATGGCTTAACCTCAGAGGATGTGGCAGAGA	1624
QY	415	yAlaProValLeuAargValThrAlaSerAapAapAapGlyserAenAlaValAlaH	435
Db	1625	GACAGCTGTGGCCCTGGTGCAGGTGTCTGACCGAGATAGAGGAGAAATGACGTGTC--	1682
QY	435	sTYserIlleuetserrGlyAenAlaArgGlyGlnPheTYrLeuAapAlaGlnThrGlyAl	455
Db	1683	-ACCTGTGTGGTGGCAGGTGATGTGCCCTTCAGGTGGCCAGGCGCAGTGAAGACGGAG	1741
QY	455	a-----LeuAapValAIserrProLeuAapTYrGluThrThrLY	468
Db	1742	TGACGACGAAGAAGATATTCCTGCGACATCAACCCCGCTTAAGCTACGAAAGGTCA	1801
QY	468	sgLufuryThrLeuAargValAargAlaGlnAapGlyGlyLysrProProLeuSerAenValse	488
Db	1802	AGACTACCACTTGAATTTGGCTGTGGATCTGTGGCAACCCCACTCTCCACACTTA	1861
QY	488	rglyLeuValThrValGlnValLeuAapIlleAenAapAapAlaProIllePheValserTh	508
Db	1862	CTCCCTC---AAGGTGACAGGTGGTGGAGTCAATGACCAACCACTGCTTCACCTCAGAG	1918
QY	508	rProPheGlnAlaThrValLeuGlnLysSerValProLeuGlyTYrLeuValLeuHISValGl	528
Db	1919	TGTACTAGAGTGGCGCTTCCCGGAAAAACAACAAGCTGTGTAAAGTGAATGCTGAGATCAC	1978
QY	528	naAlaIlleAapAlaAapAlaGlyAapAapAlaArgLeuGluTYrArgLeu-----	544
Db	1979	TGCCAGTATCTGACTCTGGCTCTAAATGCTGAAGCTGGTTTACTCTCGAGAGCTGAGCC	2038

QY	545	-AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyrPileSerVal	564
Db	2039	GGCTCTTAAGGGC-----CTCTTCAACCACTTCCACCGAGACTCGAAGATCCAGGT	2085
QY	564	AlaIaIaGluLeuAspArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAs	584
Db	2090	GAAGACATCTCTGGATCCGGAAACAGCGGAGAGCTATAGATTGAAGGTGGTGGACGTGA	2144
QY	584	PhiGlyThrProAlaLeuThrAlaSerAlaSerValThrValLeuAspValAs	604
Db	2150	CCGGGGGACGTCTTAGCTCCAGGGCAGACCACTGTCTTGTCAATGATGCTGAC	2200
QY	604	naSpAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaI	624
Db	2210	TGACATATACCCCAA-----	2225
QY	624	avaIaGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh	644
Db	2225	-----	2225
QY	644	rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGl	664
Db	2225	-----	2225
QY	664	yGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluIuArgGlnTyrValLeuAl	684
Db	2225	-----	2225
QY	684	avaIThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValaAsnValThrAs	704
Db	2225	-----	2225
QY	704	paIaAsnThrHisArgProValaPheGlnSerSerHisTyrThrValaAsnValaAsnGluAs	724
Db	2226	-----TTTATGCTAGAGGCTACCACTTCTCAGATGATGGAGAA	2263
QY	724	paArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs	744
Db	2264	CATGCAAGCACTGAGTCCAGTGGGACATGATGATCTGATGATGAGACAAAGGGGAGAA	2322
QY	744	naIaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspIaAspTh	764
Db	2324	TGCCCAAGTGGAGCTCTCAGAGGAGCAAGCAACGCTGATCTGTATCCAGAAATGGAC	2383
QY	764	rgIyAlaValThrThrGlnAlaGluLeuAspTyrGluAspIleValSerTyrThrLeuAl	784
Db	2384	AGGCACCATCTTATCCAGCTGAGCTTGTATCGAAGCAAAAGACCTACACTTCCA	2443
QY	784	alleThrIlaArgAspAsnGlyIleProGlnIlySerAspThrThrTyrIleuGluIleLe	804
Db	2444	GCTGAAGCAATGATGGTGGCTGCCACCTCGCTCAGCTTACGTGGTGCACACATCAA	2503
QY	804	uValaAspAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnIlySerVa	824
Db	2504	TGTGTGTGACAGAAATGACAGCAAGCACTTATATCACTGCCCTTCT-----AACACCTC	2557
QY	824	IlyrGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh	840
Db	2558	TCACAAGCTGTGACCCCCAGCACAGCTTGTGTGAAGCGGTGACCGAGGAGCGACGA	2617
QY	840	raSpArgAspSerGlyLeuAsnGlyArgValaPheTyrThrPheGlnIlyGlyIaAspArgI	860
Db	2618	GGACTTGTGACTGTGTGTCAATGCGGACGTGATCAACGACTTCAGATGCGAACCTCT-	2675
QY	860	yaSpGlyAspPheIleValGluSerThrSerGlyIleValaArgThrLeuArgArgLeuAs	880
Db	2676	-TATGACTCTTCCAGATGGGTGCACATTCAGGTGCATCAACCTGGAGAAAGAGATTGA	2734
QY	880	paArgGluAsnValaIaGlnTyrValLeuArgAlaTyrAlaValaAspIlySerGlyMetProPr	900
Db	2735	GGGGCGCCACATGAGGGCTACCGCCCTGTGGTGAAGGTCAAGTACCGCGGCAAGCCCC	2794
QY	900	caIaArgThrProMetGluValThrValIThrValLeuAspValaAsnAspAsnProProVa	920

```

Db      2795 ACCGTATGCAACGCTTGTGCTTATGTCATGATGACCTGTGGCAACGCAAGCT 2854
Qy      920 1PhegluIn-----AsgluPheaspValPheValGluGluAsnSerProIleG 937
Db      2855 GCTGGAGACCTCTCTGGGCCCAAGCCTGTGAC-----ACGGCGCTGGA 2896
Qy      937 yLeuAlaValAlaArgValThraAlaThrAspProAspGluGlyThraAsnAlaGln----- 955
Db      2897 TATTGACATT-----GCTGGGATCCAGATATAGAGCGCTCCAGACAGCGCTGG 2944
Qy      956 ----11MetCTyGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePh 974
Db      2945 CAACATCTCTTGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2996
Qy      974 eSerGlyGluLeuThrAlaLeuVal----- 982
Db      2997 -----GCCCTGGCGGTTCTTGTGGCTTACTGACAGACGCGGAGGCCAAAGTGGTTA 3049
Qy      983 -----AspLeuAspTyrgluAspArgProGluTyrgluValle 994
Db      3050 CCAAGCTGTGAAGAAGACCAAGACCTG-----TATGCCCCCAAGCCAGTGGCAAGC 3106
Qy      994 uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014
Db      3107 CTCGAGGAGAAACAAAGCAAGGCAAGAGCAAGACAGTCCCAAGCCGCTGAAGCCAGT 3166
Qy      1014 uAspArgAsnAspAsnProProValLeuGlyAsnPhgGluLeuPheAsnAsnTyrgVa 1034
Db      3167 GGAAGACAGAGATAGAGCCCGGCTGCAAGAGTCCCTCAAGTTC----- 3209
Qy      1034 1ThraAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054
Db      3210 ----AACTGATGATGAGATGCGCTCTGGGAGACAT-----CCCGGACATC-----CACCTGCC 3256
Qy      1054 cAspIleSerAspSerLeuThrTyrgSerPheGluArgGlyAsnGluLeuSerLeuValle 1074
Db      3257 C----- 3257
Qy      1074 uLeuAsnAlaSerThrGlyGluLeuTygLeuSerArgAlaLeuAsnAsnArgProle 1094
Db      3258 -CTCAACTACCCACCAAGCAGCAGCCTGACCTGAGCGGCGGCACATGCTCTTAATCTCCACT 3316
Qy      1094 uGluAlaIleLeuSerValLeuValSerAspGlyVal-----HisSerValThraI 1111
Db      3317 GCCTTCATCCAGCTGACGCCCACTGACCTCAAGTCCCAAGAGCAGCAGGTGATCA 3376
Qy      1111 agln-CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131
Db      3377 GGAAGCTGCAACCTGCA-----ACACATTGC----- 3402
Qy      1131 euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeug 1147
Db      3403 -----TGGGCAACCGGGGACACACAGCTCCACGGGCTCTAGACAT----- 3441
Qy      1147 TyLeuPheIleGlnAlaValAlaAlaAlaThrLeuAlaThrProProAspHisValValP 1167
Db      3442 -----ACTCCGACTACAGTACCGCACCAACCCCCCA----- 3474
Qy      1167 heAsnVal-GlnAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186
Db      3475 --AATATCCCAAGCAGAGATTACTACCGCGGCTCACTTC----- 3517
Qy      1187 ValGlyGlnProProGlyProGlyGlyGlyProProPheLeuProSerGlu 1203
Db      3518 ---GGCCACCAAGCAGCGCCAGAGCTGCAAGACCC---ATCCACAGACAG 3562

```

RESULT 6  
 US-08-453-695A-102  
 ; Sequence 102, Application US/08453695A  
 ; Patent No. 5708143  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro

```

; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta B.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; REFEATURE:
; NAME/KEY: CDS
; LOCATION: 495..4103
; US-08-453-695A-102

Alignment Scores:
Score: No.: 3,256-46 Length: 4650
Percent Similarity: 881.50 Matches: 343
Best Local Similarity: 38.23% Conservative: 146
Query Match: 26.82% Mismatches: 452
DB: 5.67% Indels: 339
Gaps: 46

US-09-916-849A-3 (1-2923) x US-08-453-695A-102 (1-4650)
Qy      3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeu 22
Db      513 AGCCCAAGCCCGTGGGGGCAAGCGGCTACTGCGCTTCATGCTGTACAGCTGCTGCTC 572
Qy      23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42
Db      573 CTGCTGCTCTCAATCCCGAG-----CAAGCCCACTGGGTAGTGA 613
Qy      43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrgluCysAspSerSerAla 62
Db      614 CAAGTGGCGGAGAGAACACCAACCAACCTT---CATTTG-----AGCTTGGC 661
Qy      63 SerAsnLeuTyrgluTyrgThrSerArgCysArgAspAlaGlyThrgluLeuThrGlyHis 82
Db      662 AGCCGACTATGTTT-----TCCAGATGTGGGCACTGTCAAGCTAGAGGTGGTGC 715
Qy      83 LeuValPro-----HisHisAspGlyLeuArgVal 92
Db      716 CCGTACCTTGGCTGATGAGCAAGACAGGTACATTTTCACCAACCGACAGCTC----- 769
Qy      93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
Db      770 -----CATGACCGTGAAGGGGCTCCGTGAATGCAAGAACAGCT 808

```

QY 111 ProTPSerCyArgLeuLeuGlyIleGlyGlyHisLeuSerProGlnGlyLysLeuThr 130  
 DB 809 CCTGTGATATCCCTGCATCTCTGAGTTGAGGT----- 841  
 QY 131 LeuProGlnGluHisProCySerLeuValProArgLeuArgCyGlnSerCySerLeu 150  
 DB 842 -----ATCTATCACAGACCTCGGCGAAGATG----- 868  
 QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGlyArgSerProGlnGlySerLeuGlyArg 170  
 DB 869 ---GAGCCCCCGGCTGCTAGAGGCGCAATAGAGT----- 901  
 QY 171 ArgGlyArg-AsnValAsnThrAlaProGlnPheGlnProProSerTyGlnAlaThrVa 190  
 DB 902 ACAAGACATCAATGACACACA---CCCACTTCGCTCCACAGATCATCATCTGCGCAT 958  
 QY 190 LProGlnGlnGlnProAlaGlyThr-----ProValAlaSerLeuArgAla 206  
 DB 959 CCTGAGAACCAACATCGGCTCATCTCCCATCCGCTGCGCTCA----- 1007  
 QY 206 AspProArgGlnGlyGlyAlaGlyArgLeuGlyThrMetAspAlaLeuPheAspSe 226  
 DB 1008 -GACCGTATGCTGGTCCCAACGATGGGATCATATGAGCTGAGGTCGAGGAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlnG 246  
 DB 1067 GGAGGAGAGCA-----CCACGCTCATTTGATGGGCA 1102  
 QY 246 uLeuAspArgGlnThrLysSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
 DB 1103 CCTGACCTGATGAGCCCTGGAGACTCCTATGACCTCAACATCAAGGTCAGATGGCGGAG 1162  
 QY 266 LProArgArgSerAlaLeuAlaThrLeuThrIleuValThrAspThrAspHisAs 286  
 DB 1163 CCCCCACGCGCAGAGTGGCCCTGCTGGGTGTCACCGGTGTCACACCAATGACCAAGCC 1222  
 QY 286 PProValPheGlnGlnGlnGlyThrLysLeuSerLeuArgGluAsnLeuGlnGlyTy 306  
 DB 1223 CCCCAGTTTGAAGCGCCCTCTATGAGCGCCGAACTATCTGAGAAATGACCCCATGGGCA 1282  
 QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTy 326  
 DB 1283 CTCGGTATCCAGGTGAGGAGCCAAATGATCATCAAGTGCATGACGAATGATGA 1342  
 QY 326 rArgLeuLeuGlnGlySerGlySerProSerGluVal-----PheGlnIleAs 343  
 DB 1343 CACATTCACACAG-----GCCGCCGAAGTTGTGAGGCGTCTTCTTCACCTGA 1390  
 QY 343 PProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGluValGlnSerTy 363  
 DB 1391 CAGGAACACTGACCTTATCATCTGTCAGGCGCCGGTGAACCGTGAAGACCTTAAGCACCT 1450  
 QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 DB 1451 GCGCTTCACAGTCTTGCTTGAAGGACCGAGCCAAAC-----CCCAAGATGCCCCGTC 1504  
 QY 383 AlaValPheLeuSerValGluAspAspAspAspAlaProGlnPhe----- 399  
 DB 1505 CCAGGTGGTGTGACCGTGAAGGACATGATGCAATGCCCCACCATTTGAGATCCGGGG 1564  
 QY 400 -----SerGlyArgTyValValGlnAlaArgGluAspValThrProG 415  
 DB 1565 CATGAGGCTAGTACTCATCAAGATGGATGGCTTAACATCTCAAGAGATGGCGAGAGA 1624  
 QY 415 ValAspProValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHis 435  
 DB 1625 GACAGCTGTGGCCGTGAGGAGGTGTCTGACCGAATAGAGGAGAGATGACGAGTCTC-- 1682  
 QY 435 rTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyLeuAspAlaGlnThrGlyAl 455  
 DB 1683 -ACCTGTGTGTGAGGATGTGTCCTTCCAGTGTGCGCAGGCACTGAGACAGGCGAG 1741  
 QY 455 a-----LeuAspValAlaSerProLeuAspTyGlnThrThrLys 468

DB 1742 TGACAGCAAGAGAGATATTCCTGACAGCTACACACCCCGCTAGACTACGAAAGTCAA 1801  
 QY 468 sGlnTyThrLeuArgValArgAlaGlnAspGlyLysArgProProLeuSerAsnValSe 488  
 DB 1802 AGACTACACCATGATGATTTGGGCTGTGGACTCTGGCAACCCCACTCTCCAGACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAspAlaProIlePheValSerTh 508  
 DB 1862 CTCCTC---AAGGTGAGAGGTGTGAGCTGTAATGACAAACCACTGCTTACACTCAAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlyTyTyLeuValLeuHisValG 528  
 DB 1919 TGTCATGAGGTGCGCTTCGCCGAAACAAACAGCTGTGTAAGATGATTCGATGATC 1978  
 QY 528 AlaIleAspAlaAspAlaGlyAspAspAlaArgLeuGlyTyTyArgLeu----- 544  
 DB 1979 TGCCAGTATCTGATCTGCTGCTTAATGCTAGAGTGTTAATCTCTGAGACCTGAGCC 2038  
 QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyThrGlyTyTyP 564  
 DB 2039 GGCTGCTAAGGC-----CTTTCACCATCTCAACCGAGCTGAGAGATCGAGT 2089  
 QY 564 LAlaAlaGlnLeuAspArgGlnGluValAspPheTySerPheGlyValGluAlaArgAs 584  
 DB 2090 GAAGACATCTCTGATCGGGAACGCGGAGAGCTATGATGGAAGGTGTGCGAGCTGA 2149  
 QY 584 PHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 DB 2150 CCGGGGCAAGTCTTACCTTCAGGGCAGACCACTGCTCTTCAATGTCTGAGCTGCA 2209  
 QY 604 nAspAsnAsnProThrPheThrGlnProGlyTyThrValArgLeuAsnGluAspAla 624  
 DB 2210 TGACATGACCCCAA----- 2225  
 QY 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 DB 2225 ----- 2225  
 QY 644 rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGly 664  
 DB 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyTyLeuGlnGluGlnTyValLeuAl 684  
 DB 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704  
 DB 2225 ----- 2225  
 QY 704 PAlaAsnThrHisArgProValPheGlnSerSerHisTyThrValAsnValAsnGluAs 724  
 DB 2226 -----TTTATGCTGAGTGGCTTACAACTTCAAGTATGAGAA 2253  
 QY 724 rArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 DB 2264 CATGCCAGCACTGAGTGCAGTGGCGCATGTGATCTGATGATGAGAACAGGGGAGAA 2323  
 QY 744 nAlaArgIleThrTyThrMetGlnAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 DB 2324 TGCCAGATGACGCTCACTGAGAGCAGAACACGCTGACTTGTATCCAGAAATGGCAC 2383  
 QY 764 rGlyAlaValAlaThrThrGlnAlaGlnLeuAspTyGluAspGlnValSerTyThrThrAl 784  
 DB 2384 AGGACATATCTTATCCAGCTGAGCTTGTATCGAGGCAACAAAGCACTTACACTTCCA 2443  
 QY 784 aIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyTyLeuGluIle 804  
 DB 2444 GCTGAAGCGAGTGAATGTGTGCTCCACTCTCACTTACGTTGCTGTCAACATCAA 2503  
 QY 804 uValAsnAspValAspAspAsnAlaProGlnPheLeuArgAspSerTyGlnGlySerVa 824

```

Db      2504 TGTGTCGACGAGATGACAGCAGCAGCCATATACCTGCCCCCTTCT-----AACCTTC 2557
Qy      824 1TyrGluAspValProPhePheThr-----SerValLeuGlnIleSerAlaTh 840
      2558 TCACAAGCTGCTACACCCCGACAGCAGCTGTGTGAGAGAGAGCTGACGAGGTGACACCA 2617
Qy      840 rAapAgaBpSerGlyLeuBnGlyAryValPheTyThrPheGlnGlyGlyAspApgl 860
      2618 GAGCTTGTACTGCTGTCTCATGATCCGAGCATCTCAAGCATCTGACGAGGTGACACCT-- 2675
Qy      860 yAapGlyAapPheIleValGluSerThrSerGlyIleValAryThrLeuAryLeuAs 880
      2676 -TATGAGCTTTCAGATGTGGTCACTTATGATGATGATGATGATGATGATGATGATGAT 2734
Qy      880 pArGluAsnValAlaGlnTyValLeuAryAlaTyAlaValAspLysGlyMetProP 900
      2735 GCGGCGCCACAGAGGCTACACGCGCTGATGAGAGTCAAGAGACCGGCGCAAGCCCC 2794
Qy      900 cAlaAryThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVa 920
      2795 ACCGTATGACAGACCCCTTGTGCTCATTTATGTCATGATGATGATGATGATGATGATGAT 2854
Qy      920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937
      2855 GCTGAGAGACCTCTCTGCGGACAGCCTGAC-----AGCGCGCTGGA 2896
Qy      937 yLeuAlaValAlaAryValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955
      2897 TATTGACATT-----GCTGGGATTCAGATATGAGAGCGCTCCAGACGAGGTGG 2944
Qy      956 ----lIeMetTyrglnIleValGluGlyAsnIleProGluValPheGlnLeuAspIleP 974
      2945 CAACATTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2996
Qy      974 eSerGlyGluLeuThrAlaLeuVal----- 982
      2997 -----GCCCTGCGGCTTCTTGTGCGCTACTGACAGCAGCGGAGCGCAAAAGTGTTA 3049
Qy      983 -----AspLeuAspTyrglyLeuAspArgProGluTyValLe 994
      3050 CCAAGCTGTATGAAGAGAGACCAAGACCTG-----TATGCCCGCAAGCGCTGAGCAAGC 3106
Qy      994 vAlaIleGlnAlaThrSerAlaProLeuValSerAryAlaThrValHisValAryLeuLe 1014
      3107 CTCCAAGGAGAAACAAAGCAAGGCAAGAGAGCAAGATCCCAAGCGCTGAGGCAAGT 3166
Qy      1014 uAspAryAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAryTyVa 1034
      3167 GAGAGAGAGAGATGAGCGCGGCTGCAAGAGTCCCTCAAGTTC----- 3209
Qy      1034 lThrAsnArySerSerSerPhePheProGlyAlaAlaIleGlyAryValProAlaHisAapPr 1054
      3210 ----AACCTGATGAGCATGCGCTGGGAGACAT-----CCCGGCATC-----CACTGCC 3256
Qy      1054 cAapIleSerArySerIleuThrTyTrSerPheGluAryGlyAsnIleuSerLeuValLe 1074
      3257 C----- 3257
Qy      1074 uLeuAsnAlaSerThrGlyGluLeuLysLeuSerAryAlaLeuAsnAsnAryProLe 1094
      3258 -CTCAACTACCCACACGAGCAGCCTGACCTGGCGCGCACATGAGCTTAATCCCACT 3316
Qy      1094 uGluAlaAlaIleMetSerValLeuValSerApglyVal-----HisSerValThrAl 1111
      3317 GCTTCCATTCAGCTGACGCGCCAGTCACCTCAAGGCTCCAAAGAGACGAGTGTATCA 3376
Qy      1111 aGln-CyAlaLeuAryValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131
      3377 GGAACCTGACACTTGAA-----ACACATTCG----- 3402
Qy      1131 eAryGluGluAspMetSerPro-----GluAryPheLeuSerProLeuLeuG 1147
      3403 -----TGGGCAACCGGAGACACCAAGCTCAAGCGGCTCTGAGCAGT----- 3441

```

```

Qy      1147 lYLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrProProAspHisValValP 1167
Db      3442 -----ACTCGACTACGACTACCGACCAACCCCGCA----- 3474
Qy      1167 hAenVal-GlnAryAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186
      3475 --AATACCCAGACAGCAGTACCTCACCGCGCGCTCACTTCTC----- 3517
Qy      1187 ValGlyGlnProProGlyProGlyGlyGlyProPheLeuProSerGlu 1203
      3518 ---GCCCAACGACGAGCGCCAGAGCTGACAGACCC---ATCCACGACAG 3562

RESULT 7
US-08-268-161A-102
; Sequence 102, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Photocatheter Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; ADDRESSER: Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,317
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 495..4103
; US-08-268-161A-102

Alignment Scores:
Pred. No.: 3, 25e-46 Length: 4650
Score: 881.50 Matches: 146
Percent Similarity: 38.23% Conservative: 146
Best Local Similarity: 26.82% Mismatches: 452
Query Match: 5.67% Indels: 339
DB: 1 Gaps: 46

US-09-916-849A-3 (1-2923) x US-08-268-161A-102 (1-4650)
Qy      3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeuLeu 22
Db      513 AGCCAGGCGCTGGGGGAGACGAGCTACGCTGCGCTCATGCTCTACACCTGCTGCTC 572
Qy      23 LeuLeuLeuProProProLeuLeuGlyAapGlnValGlyProCyBarSerLeuGlySer 42

```



Db 573 CTCTGGCTCCATCCCGAGG-----CCAGCCACTGGGGTAGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCyProSerSerAla 62  
 Db 614 CAGAGTCCCGGAGGAACGCCCAACCTCT--CATGG-----GAGCTCGC 661  
 QY 63 SerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82  
 Db 662 AGCCGACTATAGTTT-----TCCAGATGTGGGCACTGTACAGCTAGAGGTGGGTGC 715  
 QY 83 LeuValPro-----HisHisAspGlyLeuAspVal 92  
 Db 716 CCGGTACTTTCGGGTGATGGCAAGACAGTGAATTTCACCAACGAGACTTC----- 769  
 QY 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110  
 Db 770 -----CATGACCTGTAGAGGGCTCTCGTAATGCAAGAACAGCT 808  
 QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGluGlyIleThr 130  
 Db 809 CCTGTGATCCTCGCATCCCTGAGTTTGAAGT----- 841  
 QY 131 LeuProGluGluIleProCysLeuValAspAlaProArgLeuArgCysGlnSerCysIleVal 150  
 Db 842 -----ATCTATCACAGACTCTGTGACAGAAATGC----- 868  
 QY 151 AlaGluAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyIleArg 170  
 Db 869 --GAGCCCGGCTGTAGAGGCCAGATGAAGT----- 901  
 QY 171 ArgGlyArg-AspValAsnThrAlaProGluPheGlnProProSerTrpGluAlaThrVal 190  
 Db 902 ACAAGACATCAATGACACACACAC--CCCACTTCGCTCACAGTCACTCTGGCAT 958  
 QY 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206  
 Db 959 CCTGTAGAACACCAACATCGGCTCACTTCCCATCCGCTGCTTCA----- 1007  
 QY 206 eAspProAspGluGlyGluAlaGlyArgLeuGluThrMetCysAspAlaLeuPheAspSe 226  
 Db 1008 -GACCGTGAATGCTGTGCTCCCAAGGTGTGATCTTATGAGCTCAGGTGCAGAGACCA 1066  
 QY 226 TrpSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 Db 1067 GGAGGAGAACCA-----CCAGCTCATTTGTGAGGCA 1102  
 QY 246 uLeuAspArgGluThrThrSerThrHisValPheArgValThrAlaGluAspHisGlyPhe 266  
 Db 1103 CCTGACCCGTGACCGCTGGAGCTCTATATGACTCACCATCAAGGTGCAGAGTGGCGCAG 1162  
 QY 266 LProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
 Db 1163 CCCCCCAGCGGCAACAGATGCTCTGTGCTGTACACCGTCTTGACACCAATACACACG 1222  
 QY 286 pProValPheGluGlnGluGluThrGlySerLeuAspArgGluAsnLeuGluValIleTrp 306  
 Db 1223 CCCCAGATTTGACCGGCTCTATATGAGCGCGCAACTATCTGAGAAATAGCCCCATAGGCCA 1282  
 QY 306 rGluValIleuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeu 326  
 Db 1283 CTGGTCACTCAGGTGAAGCCAAATGACTCAAGACCAAGGTGCCAAATGCAAAATTA 1342  
 QY 326 TrpLeuLeuGluGlySerGlyIleSerProSerGluVal-----PheGluIleAs 343  
 Db 1343 CACATTCACACAG-----GGCCCGAAGTTGTAGGCGCTTCTTGACTGGA 1390  
 QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTrp 363  
 Db 1391 CAGAAACCTGAGCTTATCACTGTTCAGGGCCGGGTGAGCCGTGAGAGACTTAAGCACT 1450  
 QY 363 rGluLeuThrValGluIleAspArgGluGlyArgAspProGlyProArgSerThrThrAl 383

Db 1451 GCGCTTCTGAGTCTTGTATAGACCGAGCAACAC-----CCCAAGATGCCCGTGC 1504  
 QY 383 AlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPhe----- 399  
 Db 1505 CAGAGTGTGTGACCGTAAAGACATGTATATGACATATGCCCCACATTGAGATCCGGGG 1564  
 QY 400 -----SerGluValArgTrpValValGluValArgGluAspValThrProGlu 415  
 Db 1565 CATAGGCTTAGTACTCATCAAGATGGGATGGCTAACATCTCAGAGGATGTGCAGAGGA 1624  
 QY 415 ValProValIleuAspValThrAlaSerAspArgAspGlySerAsnAlaValHis 435  
 Db 1625 GACAGCTGTGGCCCTGTGTCCAGGTGTCTGACCGAGATGAGGAGAAATGCACTGTCC-- 1682  
 QY 435 rTrpSerIleMetSerGlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGGCAGGTATGTGCTTCCAGCTGCGGCCAGGCCAGGAGACAGGCGAG 1741  
 QY 455 a-----LeuAspValIleSerProLeuAspTrpGluThrThrIle 468  
 Db 1742 TGACAGCAAGAAAGTATTTCTGCAAGCTACACCCCGCTAGACTACAGAAAGTCA 1801  
 QY 468 rGluTrpThrLeuArgValArgAlaGlnAspGlyIleArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCAATTGAGATTTGGCTGTGCACTCTGGCAACCCCACTCTCCAGCACTAA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAlaProIlePheValSerThr 508  
 Db 1862 CTCCTCT--AAGGTGCAAGTGTGTGACGTCAATATACAGCACTGTCTTACCTACAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTrpLeuValLeuHisValGlu 528  
 Db 1919 TGTCACTAGAGTGTGCTTCCCGGAAACAAACAGCTGTGAAGATGTGAGATCAC 1978  
 QY 528 ValIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTrpArgLeu----- 544  
 Db 1979 TCCAGTATGTGTGACTGTGCTTATGTCTAGGTGTGTTACTCTGTGAGCTGAGCC 2038  
 QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTrpIleSerVal 564  
 Db 2039 GGTCTGTAAGGC-----CTCTTACCATCTCACCCAGACTGAGAGATCCAGGT 2089  
 QY 564 ValAlaGluLeuAspArgGluGluValAspPheTrpSerPheGlyValGluAlaArgAs 584  
 Db 2090 GAAGCATCTCTGATTCGGGAAACAGCGGAGCATATGATGTAAGGTGTGGCTGCACTGA 2149  
 QY 584 PheGlyThrProAlaLeuThrAlaSerAlaSerValIleValIleuAspValAs 604  
 Db 2150 CCGGAGCTCTTACCTTCCAGGCAACAGCACTGTCTTGCATATGTGTGACTGCA 2209  
 QY 604 AspAsnAspProThrPheThrGlnProGluTrpValArgLeuAsnGluAspAlaIle 624  
 Db 2210 TGACATATGACCCAAA----- 2225  
 QY 624 ValGlyThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleThr 644  
 Db 2225 ----- 2225  
 QY 644 rTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGly 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTrpTrpLeuGluArgGlnTrpValIleuAl 684  
 Db 2225 ----- 2225  
 QY 684 ValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValIleAsnValThrAs 704  
 Db 2225 ----- 2225  
 QY 704 ValAsnThrHisArgProValPheGlnSerSerHisTrpThrValAsnValAsnGluAs 724  
 Db 2226 -----TTTATGCTAGTGTGCTTACCACTTCTCAGTATGAGAA 2263



NAME/KEY: CDS  
 LOCATION: 495..4103  
 US-08-453-702A-102

Alignment Scores:

Prod. No.:	3,25e-46	Length:	4650
Score:	881.50	Matches:	343
Percent Similarity:	38.23%	Conservative:	146
Best Local Similarity:	26.82%	Mismatches:	452
Query Match:	5.67%	Indels:	339
DB:	2	Gaps:	46

US-09-916-849A-3 (1-2923) x US-08-453-702A-102 (1-4650)

```

QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeuLeu 22
DB 513 AGCCGAGCCCTGGGGGGGCAAGCGCTACTGCTCCCTCCATGCTGTCAGCATGCTGCTC 572
QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysValSerLeuGlySer 42
DB 573 CTGCTGCTCCATCCCAAG-----CCAGCCCACTGGGAGTGTGA 613
QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetClyThrLeuGlyProSerSerAla 62
DB 614 CAGAGTCCCGAGAGACAGCCCAACCCCT--CATTG-----GAGCCTGCG 661
QY 63 SerAsnLeuThrLeuTyThrSerArgCysArgAspAlaGlyThrGlnLeuThrGlyHis 82
DB 662 AGCCGACTATGTTTT-----TCCAGATGTGGGCACTGTACAGTACAGTACAGTGGTGC 715
QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92
DB 716 CCGGTACTTTCGGGTGATGCAAGAGGTGACATTTTACACACGAGACCTC----- 769
QY 93 TrpCysProGluSerGluAlaHisGlyPro-----LeuProProAlaProGluGlyCys 110
DB 770 -----CATCAGCCGTAGAGGGCTCCGTGATGCAAGACCAAGCT 808
QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleuSerProGluGlyLeuThr 130
DB 809 CCTGTGTGATCCTGTGATCCTGTGAGTTGAGGT----- 841
QY 131 LeuProGluGlnHisProCysLeuValAlaProArgLeuArgCysGlnSerCysValLeu 150
DB 842 -----ATCTATCAGACCTCGTGCAAGATGC----- 868
QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyArg 170
DB 869 ---GAGCCCCGGCTGTAGAGGCCAGATGAGGT----- 901
QY 171 ArgGlyArg-AsnValAsnThrAlaProGlnPheGlnProProSerTyrglnAlaThrVal 190
DB 902 ACAAGCATCATGACACACACA-----CCCACTTGGCTCACAGTACATCTGGCCAT 958
QY 190 LProGluGlnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206
DB 959 CCTGTGAAACCAACATCGGCTCACTTCCCATCCCGCTGCTTCA----- 1007
QY 206 eAspProAspGluGlyGlnAlaGlyArgLeuGlyThrThrMetCAspAlaLeuPheAspSe 226
DB 1008 -GACCGTATGCTGTGCTCCCAAGGTGTGGCATCTTATGAGCTCAGGTGCGAGAGACA 1066
QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGluG 246
DB 1067 GGAGAGAGAGCAA-----CCAGACTCATTTGTATGGGCAA 1102
QY 246 uLeuAspArgGluThrIlySerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266
DB 1103 CCGGACCGGTGACGCTGGGACTCTTAATGACCTCACCATCAGTACAGTGCAGATGCGCGCAG 1162
QY 266 cProArgSerSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286
DB 1163 CCCCCACGGGCAACAGTGCCTGTGCTGTATCAGCGTGTGACCTTGAACCAATGACACGC 1222
  
```

```

QY 286 pProValPheGluGlnGluGlyThrIleGlySerLeuArgGluAsnLeuGluValGlyTy 306
DB 1223 CCCCAGTTTACGGGCGCTCTTATGAGCCGGAATCTTGTGAAATAGCCCCATGAGCCA 1282
QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTy 326
DB 1283 CTGCGTATCAGGTGAGGAGCCAAATGACTCAGACCAAGGTGCCAATGAGAAATGCAAT 1342
QY 326 rArgLeuLeuGluGlySerGlySerProSerGlyVal-----PheGluLeuAs 343
DB 1343 CACATTCACACAG-----GCGCCCAAGTTGTGAGGCGTCTTTCGACCTGGA 1390
QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGlySerTy 363
DB 1391 CAGAAACATGACTTATACCTTACGTTCAGGCCCCGTGACCGGTGACCACTTACACACCT 1450
QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383
DB 1451 GCGCTTCTCAGTGTGCTTGAAGGACCGAGGACCAAC-----CCCAAGATGCCCGTGC 1504
QY 383 aAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPhe----- 399
DB 1505 CCAAGTGTGTGACCGTGAAGACATGATACATATGCCCCACCATTTGAGATCCGGGG 1564
QY 400 -----SerGluValArgTyThrValGlnValArgGluAspValThrProGl 415
DB 1565 CATAGGGCTATGATCTCATCAAGATGAGTGCATACATCTCAGAGAGATGTGCAAGAGA 1624
QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValAlaHis 435
DB 1625 GACAGCTGTGGCCCTGTGTGACAGTGTCTGACCAAGATGAGGAGAGATGACGCTGTC-- 1682
QY 435 rTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyThrLeuAspAlaGlnThrGlyAl 455
DB 1683 -ACGTGTGTGTGAGGAGTGTGTCCTTCACTGCGGCCAGGCCAGTGAAGACAGGCGAG 1741
QY 455 a-----LeuAspValAlaSerProLeuAspTyrgluThrThrTy 468
DB 1742 TGACAGCAAGAGAGATATTTCTTCTCAGACTACACCCCGCTAGACTAGAGAAAGTCAA 1801
QY 468 eGluTyThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488
DB 1802 AGACTTACACCATTTGATTTGTGGCTGTGGACTGTGGCAACCCCACTCTCCACACTTAA 1861
QY 488 rGlyLeuValThrValGlnValLeuAspIleAspAspAsnAlaProIlePheValSerTh 508
DB 1862 CTCCCTC--AAGGTGACAGGTGTGAGCTCATATGACACGCACTGTCTTCACTCAGAG 1918
QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyThrLeuValLeuHisGly 528
DB 1919 TGTCACTAGAGTGGCGCTTCCCGGAAACAAACAGCGTGTGAAAGTGTGAGATTCAC 1978
QY 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTyArgLeu----- 544
DB 1979 TCCAGATGATGTGACTGTGGCTTAATGCTGATGCTGTGTTACTCTGAGACCTGAGCC 2038
QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyTrpIleSerVa 564
DB 2039 GCGTGTCAAGGCG-----CTCTTCACTTCAACCCAGACTGAGAGATCCAGGT 2089
QY 564 lAlaAlaGluLeuAspArgGluGluValAlaAspPheTySerPheGlyValGluAlaArgAs 584
DB 2090 GAAAGCATCTGTGATCGGAAACAGCGGAGAGCATATGATTGAAGTGTGTGCGACGTGA 2149
QY 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604
DB 2150 CCGGGGACGCTTACCTTCAAGGCGACAGCACTGTCTTGTCAATGTGTGAGTGTGCA 2209
QY 604 nAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624
DB 2210 TGACAAATACCCCAAA----- 2225
  
```

[illegible]

Db	2997	-----GCCCTGGCGGTTCTTGTGGCCCTACTGCAAGACACCGGAGGCCCAAAAGTGGTTA	3049
Oy	983	-----ApLeuAaPTrYrGluaSpArGrProGluTrValle	994
Db	3050	CCAAGCTGTGAAGAGAGACCAAGAGCTTG---TATGCCCCCAAGCCAGTGGCAAGGC	3106
Oy	994	UValleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValAlaGleuLe	1014
Db	3107	CTCCAAAGGAAACAAAGCAAGAGCAAGAAAGCAAGCATCCCCCAAGCCGTGAAGCCAGT	3166
Oy	1014	UASpArGAsnaSPaSnProRoValleuGlyAsnPhgLuIleLeuPhheaSnAenTyRva	1034
Db	3167	GGAGGACAGAGATGAGGCCGGAGCTGACAGAAAGTCCCTCAAGTTC-----	3209
Oy	1034	lThnAaPArGrSerSerSerSerProGlyGlyAlaIleGlyArgValProAlaHisAaPr	1054
Db	3210	---AACCTGATGAGCATGATGCCCTGGGAGACAGT---CCCCGATTC-----CACCTGCC	3256
Oy	1054	oASpIlleSerAaPserLeuThrTySerPheGluArgGlyAsnGluLeuSerLeuValle	1074
Db	3257	C-----	3257
Oy	1074	uLeuAaAlaSerThrGlyGluLeuTybLeuSerArgAlaLeuAaSPaSnAaGrProle	1094
Db	3258	---CTCAACTACCCACCAAGAGAGCCCTGTGACTGGGGCCGCACTATGCTTCACTCCCACT	3316
Oy	1094	uGluAlaIlMeSerValleuValSerAaPglVal-----HisSerValThrAl	1111
Db	3317	GCTTTCATTCAGGTGACGCCCACTGATCACTCAAGCTCCAAAGAACCAAGGTGGTACA	3376
Oy	1111	agIn-QyValaLeuArgValThrIleIleThrAaPglMeLeuThrHisSerIleThrL	1131
Db	3377	GGACCTGCCACCTGCAA-----ACACATTGG-----	3402
Oy	1131	euaGleuGluAaPMeSerPro-----GluArgPheLeuSerProLeuGlu	1147
Db	3403	-----TGGGCACCGGGGACACCAAGCTCCAGGGCTGTAGACGT-----	3444
Oy	1147	lyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrProProAaSPHisValValAlP	1167
Db	3442	-----ACTCCAGTACACTACCTACCGCACCAACCCCCCA-----	3474
Oy	1167	heaenVal-GlnAaGAPThrAaSPAlaProGlyGlyHisIleLeuAaenValSerLeuSer	1186
Db	3475	--AAATACCCCAAGCAAGATTACTCAACCGCGGTCAACTCTTC-----	3517
Oy	1187	ValGlyGlnProProGlyProGlyGlyGlyProPheLeuProSerGlu	1203
Db	3518	---GGCCACCAAGCCAGGCCCAAGAGGTGAGAGAACCC---ATCCACAGACAG	3562
RESULT 9	US-09-099-639-102		
	Sequence 102, Application US/09099639		
	Patent No. 6262237		
	GENERAL INFORMATION:		
	APPLICANT: Suzuki, Shintaro		
	TITLE OF INVENTION: Protocadherin Materials and Methods		
	NUMBER OF SEQUENCES: 115		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &		
	CITY: Chicago		
	STATE: Illinois		
	COUNTRY: USA		
	ZIP: 60606		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: floppy disk		
	COMPUTER: IBM PC compatible		
	OPERATING SYSTEM: PC-DOS/MS-DOS		
	SOFTWARE: Patent In Release #1.0, Version #1.25		
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER: US/09/099,639		

FILING DATE: 18 JUN 1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/263,161  
 FILING DATE: 27 JUN 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Greta E. No. 626237and  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 27866/34703  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 102:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4650 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 495..4103  
 US-09-099-639-102

Alignment Scores:  
 Pred. No.: 3,25e-46 Length: 4650  
 Score: 881.50 Matches: 343  
 Percent Similarity: 38.23% Conservative: 146  
 Best Local Similarity: 26.82% Mismatches: 452  
 Query Match: 5.67% Indels: 339  
 gaps: 46

US-09-916-849a-3 (1-2923) x US-09-099-639-102 (1-4650)

QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
 DB 513 AGCCGAGGCGCTGGGGGCAAGGCTACGCTGCTCCAGTGGCTAGCACTGCTGCTC 572  
 QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCyValGserLeuGlySer 42  
 DB 573 CTGCTGGCTCCATCCCAAG-----CCAGGCACTGGGTATGTA 613  
 QY 43 ArgGlyArgGlySerGlyValAlaCyValProMetGlyTyrLeuCyValProSerSerAla 62  
 DB 614 CAAGGTCGCGGAGAACAGCCACCAACCTT---CATTTG-----GAGCTTGGC 661  
 QY 63 SerAsnLeuTyrLeuTyrThrSerArgCyValAspAlaGlyThrGlnLeuThrGlnHis 82  
 DB 662 AGCCGACTATGTTT-----TCAGATGTGGGCACTGTACAGCTAGAGTGGGTGC 715  
 QY 83 LeuValPro-----HisValAspGlyLeuVal 92  
 DB 716 CCGGTACTTGGGTGATGCAAGACAGGTGACATTTTCAACACGAGACCTC----- 769  
 QY 93 TyrCyProGlnSerLeuAlaHisLeuPro-----LeuProAlaProGlnGlyCyVal 110  
 DB 770 -----CATGACCGTGGAGGGCTCCGTGAATGCCAGAACCAAGCT 808  
 QY 111 ProTyrSerCyValGlyLeuGlyGlyLeuGlyHisLeuSerProGlnGlyValLeuThr 130  
 DB 809 CCTTGATGATCCCTGCACTCTGAGTTTGAAGT----- 841  
 QY 131 LeuProGlnGlnHisProCyValLeuValaProArgLeuArgCyGlnSerCyValLeu 150  
 DB 842 -----ATCTATCAACAGCTCGTGCAGAAATGC----- 868  
 QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGlnArgSerProGlnGlnSerLeuGlyValArg 170  
 DB 869 ---GAGCCCCCGCTGCTAGAGGGCCAGATAGAAAT----- 901  
 QY 171 ArgGlyArg-AspValAsnThrAlaProGlnPheGlnProProSerTyrGlnAlaThrVal 190  
 DB :::: ||||| ||||| ||||| ::::

DB 902 ACAAGACATCAATGACAACA-----CCCACTTGGCTGACCAAGTCACTTGGCCAT 958  
 QY 190 LProGlnAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206  
 DB 959 CCTTGAGAACCAACAATGGCTCACTTCCCATCCCGCTGCTTCA----- 1007  
 QY 206 eAspProAspGlnGlyGlnAlaGlyArgLeuGlnTyrThrMetAspAlaLeuPheAspSe 226  
 DB 1008 -GACCTGTATGCTGTGCCCAAGGTGTGCACTTGTAGCTGCAGGTGGCAGAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGln 246  
 DB 1067 GAGAGAGAACCA-----CCACAGCTATTTGATGGGCAA 1102  
 QY 246 uLeuAspArgGlnThrTyrSerThrHisValPheArgValThrAlaGlnAspHisGlyLe 266  
 DB 1103 CCGTGAACGTGAGCGGTGGACTCTCATATCACTCAACCAATGAGTGCAGATGGCGGCA 1162  
 QY 266 tProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
 DB 1163 CCCCCCAGCGCCAGAGTGCCTGCTGCTGTCAACCGTGTGACCAACATGACCAACCC 1222  
 QY 286 pProValPheGlnGlnGlnGlnTyrLeuSerLeuArgGlnAsnLeuGlnValGlyTyr 306  
 DB 1223 CCCCAAGTTTGAAGCGCTCTCTATGAGGCCCAACTATCTGAGAAATGACCCATAGGCCA 1282  
 QY 306 rGlnValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyr 326  
 DB 1283 CTGGGATCCAGGATGAGGAGCCAAATGATCAACCAAGTGCATGCAAGAAATGCAAT 1342  
 QY 326 rArgLeuLeuGlnGlySerGlySerProSerGlnVal-----PheGlnIleAs 343  
 DB 1343 CACATTTCCACAG-----GCCGCCAAGTTGTGAGGCGCTTCTTGCATGCA 1390  
 QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnValGlnSerTyr 363  
 DB 1391 CAGGAACATGGAATTATCACTTTCAGAGGCCCGGAGCCGTGAGACCTAGCAACCTT 1450  
 QY 363 rGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 DB 1451 GCGCTTTCAGTGCCTTGTGAAGACAGAGGACCAAC-----CCCAAGATGGCCGTGC 1504  
 QY 383 aAlaValPheLeuSerValGlnAspAspAsnAspAsnAlaProGlnPhe----- 399  
 DB 1505 CCAAGTGTGTGACCGTGAAGACATGATACATGATGCCCCCACCATTGATGATCGGGG 1564  
 QY 400 -----SerGlnTyrArgTyrValGlnValArgGlnAspValThrProG 415  
 DB 1565 CATTAAGGCTAGTACTCATCAAGATGGATGGCTTAACATCTCAAGATGTGGCAGAGGA 1624  
 QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValAlaHis 435  
 DB 1625 GACAGCTGTGGCCCTGTGGAGAGGTGTGACCGGATGAGGAGAAATGACGCTGTC-- 1682  
 QY 435 sTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAl 455  
 DB 1683 -ACCTGTGTGTGAGAGTGAATGTGCTTCCAGCTGGCGCCAGGCAAGTGAAGACGGCG 1741  
 QY 455 a-----LeuAspValValSerProLeuAspTyrGlnThrThrTyr 468  
 DB 1742 TACACGACGAAGAAGATTTCTCTGACAGTACACCCCGCTGATCTACAGAAAGTCAA 1801  
 QY 468 gGlnTyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerSerAsnValSe 488  
 DB 1802 AGACTTACACATTAAGATGTGGCTGTGAGCTTGGCAACCCCACTTCACAGACTAA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508  
 DB 1862 CTCCCTC---AAGGTGAGGTGTGAGCTCAATACACAGCACTGTCTTCACTCAAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlnTyrTyrLeuValLeuHisValG 528  
 DB 1919 TGTCACTAGAGTGGCTTCCCGGAAACAACAAGCTGTGAAAGTATTCGTAATCATC 1978

QY 528 nalaaleaapalaaapalagluyaapaaalaaargleugluytargleu----- 544  
 Db 1979 TGCAAGAGAGCTGATGCTGCTGCTATGCTGAGCTGTTATCTCTGAGCCCTGAGCC 2038  
 QY 545 -AlaGlyValGlyHisPhePheProPheThrIleAsnAnglyThrGlyTrpIleSerVa 564  
 Db 2039 GGGTGTCAAGGGGCTCTTACCATCTCAACCGAGCTGAGATCCAGGT 2089  
 QY 564 lAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyValGluValArgAs 584  
 Db 2090 GAAGACATCTCTGATGAGGAGACGGAGAGCTATGAGTTGAAGGTGGTGGAGCTGA 2149  
 QY 584 pHISGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGAGAGCTCTAGCTCCAGGGGACAGACAGCTGCTTGTCAATGTGAGCTGCA 2209  
 QY 604 nAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaI 624  
 Db 2210 TGACAAATGACCCCAAA----- 2225  
 QY 624 aValIGlyThrSerAlaValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 QY 644 rTyrglnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGlnTyrValLeuAl 684  
 Db 2225 ----- 2225  
 QY 684 aValIThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleValValAsnValThrAs 704  
 Db 2225 ----- 2225  
 QY 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAs 724  
 Db 2226 -----TTTATGCTGAGTGGCTACAACTTCTCAGTATGAGAA 2263  
 QY 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCCAGCACTGATGCTCAGTGGGATGATGCTGATGATGATGATGATGATGATGATG 2323  
 QY 744 nAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 Db 2324 TGCCAGAGTGCAGCTCAGTGCAGAGCAAGCAAGCTGATGTTGTTCCAGATGGCAC 2383  
 QY 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAl 784  
 Db 2384 AGGCACATCTCTTACCAAGCTGAGCTTGTATCCAGACCAAAAGCACTTACCTTCCA 2443  
 QY 784 alleThrAlaArgAsnAnglyIleProGlnIlySerAspThrThrTyrLeuGluIleLe 804  
 Db 2444 GCTGAAGGCAAGTGGATGGTGGCTGCCAAGCTGCTGAGTTACCTGGTGTGACCATCAA 2503  
 QY 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnIlySerVa 824  
 Db 2504 TGCTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2557  
 QY 824 lTyrgluAspValProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCACAAACCTGTCAGCCCGGACAGACGCTTGTGAGACAGGTCAGCCAGGTGGAGCCGA 2617  
 QY 840 rAspArgAspSerGlyLeuAsnAnglyArgValPheTyrThrPheGlnGlyGlyAspArgI 860  
 Db 2618 GAGCTTTCAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2675  
 QY 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLeuAs 880  
 Db 2676 -TATGAGCTCTTCCAGATGGGTGATGATGATGATGATGATGATGATGATGATGATG 2734

QY 880 pArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspArgGlyMetProPr 900  
 Db 2735 GGGGGCCACCAATGGGCTACACCGCTGGTGGTGAAGTCAAGTACCGGGGAGAGCCCG 2794  
 QY 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAsnProProVa 920  
 Db 2795 ACCTATGGACAGACCTTGGTGCATCTTATATGATGATGATGATGATGATGATGATG 2854  
 QY 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleGl 937  
 Db 2855 GCTGAGAGACCTTCCTGGGACAGACCTGAGC-----ACGCGCTGGA 2896  
 QY 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955  
 Db 2897 TATTCACATT-----GCTGGGATTCAGAAATATGAGCGCTCCAGACGAGCTGG 2944  
 QY 956 -----lMetTyrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIleTh 974  
 Db 2945 CAACATTCCTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2996  
 QY 974 eSerGlyGluLeuThrAlaLeuVal----- 982  
 Db 2997 -----GCCCTGGGCTTCTGTGGCTATGCAAGACGGGAGCCAAAGTGTGTTA 3049  
 QY 983 -----AspLeuAspTyrGluAspArgProGluTyrValLe 994  
 Db 3050 CCAGCGTGTGAAGAAGAGACCAAGGACCTG---TATGCCCCCAAGCCAGTGGCAAGGC 3106  
 QY 994 uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014  
 Db 3107 CTCCAAGGAAACAAAGGAAAGGCAAGAGCAAGTCCCAAGCCCGTGAAGCAAGT 3166  
 QY 1014 uAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrVa 1034  
 Db 3167 GAGAGACAGAGATGAGGCGGCTGCAAGATCCCTCAAGTTC----- 3209  
 QY 1034 lThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054  
 Db 3210 ---AACCTGATGAGGATGAGCCCTGGGAGAGT---CCCCGATC---CACCTGCC 3256  
 QY 1054 oAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLe 1074  
 Db 3257 C----- 3257  
 QY 1074 uLeuAsnAlaSerThrGlyGlyLeuLeuSerArgAlaLeuAspAsnArgProLe 1094  
 Db 3258 -CTCAACTACCAACAGAGGAGCGCTGAGCTGGGCGGCACTATGCTTAACTCCCACT 3316  
 QY 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111  
 Db 3317 GCCTTCATTCAGCTGCAAGCCCAAGTCACTCAGGCTCCAAAGAACCAAGAGTGTACA 3376  
 QY 1111 agln-CyAlaLeuArgValThrIleIleThrAspIleMetLeuThrHisSerIleThrI 1131  
 Db 3377 GAACCTGCCACCTGCA-----ACACATTCC----- 3402  
 QY 1131 euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeuG 1147  
 Db 3403 -----TGCGACCGGGGAGACACAGTCCAGCGGCTGAGACAGT----- 3441  
 QY 1147 lYleuPheIleGlnAlaValAlaIleThrLeuAlaThrProProAspHisValValAlP 1167  
 Db 3442 -----ACTCGACTACAGCTTACCGCAACCAACCCCGCA----- 3474  
 QY 1167 heAsnVal-GlnArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186  
 Db 3475 ---AATACCCAGAGACATTAATCAACCCCGCTCACTTTC----- 3517  
 QY 1187 ValGlyGlnProProGlyProGlyGlyProProPheLeuProSerGlu 1203  
 Db 3518 ---GGCCACAGCAGGCGCAGAGAGCTGCAAGACC---ATCCACAGACAG 3562

RESULT 10



PCT-US93-12588-102  
Sequence 102, Application PC/TUS9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESS: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4650 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 495..4103  
PCT-US93-12588-102  
Alignment Scores:  
Pred. No.: 3.25e-46 Length: 4650  
Score: 881.50 Matches: 343  
Percent Similarity: 38.23% Conservative: 146  
Best Local Similarity: 26.82% Mismatches: 452  
Query Match: 5.67% Indels: 339  
Gaps: 46  
US-09-916-849a-3 (1-2923) x PCT-US93-12588-102 (1-4650)  
QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
Db 513 AGCCAGGCGCTGGGGGCAACGGCTACTGCTGCCCTCCAGCTGAGCACTGCTCTC 572  
QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42  
Db 573 CTGCTGCTCATCCCAAG-----CCAGCCCACTGGGTAAGTGA 613  
QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSerSerAla 62  
Db 614 CAAGTGGCGGAGAGCAACCAACCACTCT---CATGG-----GAGCCCTGCC 661  
QY 63 SerLeuLeuThrLeuGlyThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82  
Db 662 AGCCGACTATGATGTT-----TCCAGATGTGGGCACTGTACAGCTAGAGTGCGTGC 715  
QY 83 LeuValPro-----HisHisAspGlyLeuValArgVal 92

Db 716 CCCTACCTTCGGCGTATGCAAGAGGTGACATTTTACCAACCGAGACTTC----- 769  
QY 93 TrpCysProGluSerGlnValHisIlePro-----LeuProAlaProGluGlyCys 110  
Db 770 -----CATTCAGCTGAGGGGCTCTCGTAATGCAAGAACCACT 808  
QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGluGlyIleuThr 130  
Db 809 CCCTGGTGAATCCCTGATCTCGAGGTTGAGGT----- 841  
QY 131 LeuProGluGlnHisProCysLeuValAlaProArgLeuArgCysGlnSerCysValLeu 150  
Db 842 -----ATCTATCACAGACTCTGTCAGAAATGC----- 868  
QY 151 AlaGlnAlaProGlyLeuArgAlaGlyIleArgSerProGluGluSerLeuGlyIleArg 170  
Db 869 ---GAGCCCCGGGCTGTAGAGGCCCAATGAAAGT----- 901  
QY 171 ArgIleArg-AsnValAsnThrAlaProGluPheGlnProProSerTrpGlnAlaThrVal 190  
Db 902 ACAAGACATCAATGACAGACA---CCAACTTGCTTCACAGTCATCACTGGGCACT 958  
QY 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206  
Db 959 CCCTGAGAACCAACCAATCGGCTCACTTCCCATCCGCTGCTTCA----- 1007  
QY 206 eAspProAspGluGlyIleGlnAlaGlyIleArgLeuGlyIleThrMetAspAlaLeuPheAsp 226  
Db 1008 -GACCGTGAATCTGTGGTCCCAACGATGTGGCATCTTATGACTCAGGTGCAAGAGCA 1066  
QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
Db 1067 GAGAGAGAGACA-----CCAGACTCATTTGTATGGGCAA 1102  
QY 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGly 266  
Db 1103 CCGAGACCGTGAACGCTGGAGCTCTATGACCTCACCATCAAGGTGAGATGGCGGAG 1162  
QY 266 tProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
Db 1163 CCCCCACGGCGCAACAGATGCTCTGCTGCTGTCACCGTCTTGAACCAATGACAAAGC 1222  
QY 286 pProValPheGluGlnGlnGluIleuTrpLysGluSerLeuArgIleAsnLeuGlyValGly 306  
Db 1223 CCCCAAGTTTGAAGCGCTCTATGAGGCCCAATCTATCTGAGATATGCCCCATGAGCCA 1282  
QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleuTrp 326  
Db 1283 CTGGTCATTCAGGTGAGAGGCCAATGATCTGACCAAGGTGCCAATGCAAGAAATCGAAT 1342  
QY 326 rArgLeuLeuGluGlySerGlyValLysSerProSerGluVal-----PheGluIleAs 343  
Db 1343 CACATTCACACAG-----GCGCCGAAGTTGTAGAGCGCTCTTGGACTGGA 1390  
QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTrp 363  
Db 1391 CAGGAACACTGACTATCACTGTTCAGGGGCCGGTGAACCGGAGAGCACTTAAGCACTC 1450  
QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyValAspProGlyProArgSerThrThrAl 383  
Db 1451 GCGCTTCTCAGTCTGTCTTGAAGCCGAGCACAAAC-----CCCAAGAGTCCCGTGC 1504  
QY 383 alaValaPheLeuSerValGluAspAspAsnAspAlaProGlnPhe----- 399  
Db 1505 CAGAGTGTGTGACCGTGAAGACATGATGATGATGATGATGATGATGATGATGATGATG 1564  
QY 400 -----SerGluValArgTrpValValGlnValAlaArgGluAspValThrProG 415  
Db 1565 CATAGGCTAGTATCATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1624  
QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspArgGlySerAsnAlaValHis 435

Db	1625	GACAGCTGTGGCCCTGGTGCAGGGTGTCTGACCCGAGATGAGGAGAAAGATGAGCTGTC--	1682
Oy	435	STYrSerIleMetSerGlyAenAlaArgGlyGlnPheTyrlleuAspAlaGlnThrGlyAl	455
Db	1683	-ACCTGTGTGTGACAGGTGATGTGGCCCTTCACACTGGCCAGGCCAGTGAAGACGACAG	1741
Oy	455	a-----LeuAspValIserProleuAspTyrlGlnThrIly	468
Db	1742	TGACAGCAAGAAAGATTTCTCTGCACAGCTACACACCCCGTTAACTACAGAAAGCTCA	1801
Oy	468	SGlnTyThrIleuArgValArgAlaGlnAspGlyGlyArgProleuSerAenValSe	488
Db	1802	AGACTACACCATTAAGATGTGGCTGTGGACTGTGGACACCCCACTTCACACACTAA	1861
Oy	488	rglyleuValThrValGlnValIleuAspIleAsnAspAsnAlaProIlePheValSerTh	508
Db	1862	CTCCCTC---AAGGTGCAGGTGGTGAAGTCAAGACAGACACCTGTCTCACTCAGAG	1918
Oy	508	rProPheGlnAlaThrValIleuGluSerValProleuGlyTyrlleuValIuHrValGl	528
Db	1919	TGTCACTAGAGTCGCTTCGCCGAAAAACAACAGCGCTGTGAAGTGAATGTGTGATTCAC	1978
Oy	528	nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlnTyArgLeu-----	544
Db	1979	TGCCAGTATCTGACTCTGGCTCTAAAGCTGACGTGTATTACTCTGGAGCCTGAGCC	2038
Oy	545	-AlaGlyValGlyIleAspPheProPheThrIleAsnAsnGlyThrGlyTyrlIleSerVa	564
Db	2039	GGCTGCTAAGGC-----CTCTCAACATCTCACCCGACAGCTGAGAGATCCAGT	2089
Oy	564	lAlaAlaGluLeuAspArgGluGluValAspPheTyrlSerPheGlyValGluAlaArgAs	584
Db	2090	GAAGACATCTCTGGATCGGGAACAGCGGAGAGACTATAGTTGAAGGTGGTGACGCTGA	2149
Oy	584	pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValIleuAspValAs	604
Db	2150	CCGGGGCAGTCTACCTTCACAGGGCAGACACCTGTCTTGTCAATGTCTGACCTGCAG	2209
Oy	604	nAspAsnAsnProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl	624
Db	2210	TGACATGATCAACCAAA-----	2225
Oy	624	aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh	644
Db	2225	-----	2225
Oy	644	rTyrlGlnIleThrSerGlyAenThrArgAsnArgPheSerIleThrSerGlnSerGlyGl	664
Db	2225	-----	2225
Oy	664	ylGlyLeuValSerLeuAlaLeuProleuAspTyrlIysLeuGluIuArgGlnTyValLeuAl	684
Db	2225	-----	2225
Oy	684	aValThrIaSerAspGlyTyThrArgGluAspThrAlaGlnIleValValaAenValThrAs	704
Db	2225	-----	2225
Oy	704	pAlaAsnThrAlaArgProValPheGlnSerSerHisrTyThrValaAenValaAenGluAs	724
Db	2226	-----TTTATCGTAGGTGCTCAACCTTCAGTATGAGAA	2263
Oy	724	pArgProAlaGlyTyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs	744
Db	2264	CATGCCAGACCTGAGTCCAGTGGTGACTGTGCATTGATGAGAGAAAGAGGGAGAGA	2322
Oy	744	nAlaArgIleThrTyrlPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh	764
Db	2324	TGCCAAGGAGAGCTCTGAGTGAGAGAGCAACAGGTGACTTTGTATTCAGAAATGGCAC	2383
Oy	764	rglyValaValThrThrGlnAlaGluLeuAspTyrlGluAspGlnValSerTyThrIleuAl	784
Db	2384	AGGCACCATCTAATCCAGCTGAGCTTTGATGAGAGCAACAGACCTTACACCTTCCA	2443

QY	784	alleThrAlaArgAspAsnIylIeProGlnIuYseSerPthThrTytleuGlnIle	804
Db	2444	GCTAAGACAGTGGATGTGGCGGCCCTCGCTCAGCTTACGTTGCTGTCACCACTCA	2503
QY	804	uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrgInIseVa	824
Db	2504	TGTGCTGAGAGAAATGACAACGCCACTATATCACTGACCCCTCT-----AACACCT	2557
QY	824	IlyrGluAspValProProPheThr-----SerValIeuGlnIleSerAlaTh	840
Db	2558	TCACAAACCTGTCAGCCCCAGACACGCTTGTGTGACAGCGTACGCCAGGTGGAGCCGA	2617
QY	840	rAspArgAspSerGlyLeuLeuAsnGlyArgValPheTyTrPheGlnIylGlyAspAspG	860
Db	2618	GGACTTTCGACTCTGTGTGTCAATGCCAGCGATGTATACAGCATTCGACGTGGCAACCT--	2675
QY	860	yAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAs	880
Db	2676	-PATGACTCTTCAGATGTGGTCACTTCAGTTCAGGTGCATCACTTCGAGAAAGAGAAATTA	2734
QY	880	paArgGluAsnValAlaGlnTyValLeuArgAlaTyAlaValAspArgGlyMetProP	900
Db	2735	GCGCGGCCACATGGGCTACACCGCCCTGGTGTCAAGTCAGTACGTACCGCGCAAGCCCC	2794
QY	900	calArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVa	920
Db	2795	ACGCTAATGGACACACCTGTGGTCCATCTTATATGTCATATGACATGACCTGGCAACCCGACGT	2854
QY	920	lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG	937
Db	2855	GCTGAGACCCCTCTCGGCGCACACGCTTGAC-----ACGCGCTTGA	2896
QY	937	ylLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----	955
Db	2897	TATTCACACT-----GCTGGGAGATCCAGATATATGAGCGGCTCCAGACGCTGG	2944
QY	955	---lleMetTyrgInIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePh	974
Db	2945	CAACATCTCTCTTGTGTGTGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2996
QY	974	eSerGlyGluLeuThrAlaLeuVal-----	982
Db	2997	-----GCCCTGGCGGCTTCTGTGTGCGCTACTCGACACAGCGGAGGCCAAAGTGGTTA	3049
QY	983	-----AspLeuAspTyrgIleuAspArgProGluTyrValLe	994
Db	3050	CGAGGCTGTAAAGAGAGACCAAGACCTG---TATGCCCCCAAGGCCCATGTGGCAAGGC	3106
QY	994	uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuIle	1014
Db	3107	CTCCAAAGGAAACAAAMACAAAGCCMAAMAGCAAGTCCCAAGACCCGTGAAACCACT	3166
QY	1014	uAspArgAsnAspAsnProProValIleGlyAsnPheGlnIleLeuPheAsnAsnTyVa	1034
Db	3167	GGAGACACAGATTAAGCGCGGCGCTGCAGAAATCCCTCAAGTTC-----	3209
QY	1034	IThrAsnArgSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr	1054
Db	3210	---AACCTGATGAGCGATGCGCTCGGAGAACGT---CCCGGCATC-----CACCTTCC	3256
QY	1054	caAspIleSerAspSerLeuThrTySerPheGluArgGlyAsnGluLeuSerLeuValLe	1074
Db	3257	C-----	325
QY	1074	uLeuAsnAlaSerThrGlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArgProle	1094
Db	3258	-CTCAATTAACCAACAGGACGCGCTGACCTGGGCGCGACATATCGCTTAATCCCCACT	3316
QY	1094	uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThra	1111
Db	3317	GCGTTCATTCAGGTCGACGCCCAAGTCAACCTTCAGCTTCGAAGAACACACAGGTGTACA	3376

QY 1111 agin-cysalaleuargvalthrilethethrappglumectleuthrhisertlethr 1131  
 Db 3377 GAACTTCACCTGCA-----ACACATTCG----- 3402  
 QY 1131 euargleugluapmetserpro-----gluargphenleuserproleueng 1147  
 Db 3403 -----TGGGCACTGGGGGACACACGCTCCAGGGCTGTAGAGT----- 3441  
 QY 1147 lyleupheileginalavalalaalathrleuathrproproaspheivalvalp 1167  
 Db 3442 -----ACTCCGACTACAGCTACCGACCAACCCCCCA----- 3474  
 QY 1167 healeuval-glunargapthrrapalaprogllyghistileuvalserleuser 1186  
 Db 3475 --AATACCCGAGAGAGATTACTACCGCGGTCACTCTC----- 3517  
 QY 1187 valglvglnproproglprogllygllyprophenleuproserglu 1203  
 Db 3518 ---GGCACACGACGACCCAGAGAGCTGACAGACC---ATCCAGACAG 3562

RESULT 11  
 PCT-US95-08071-102  
 ; Sequence 102, Application PC/TUS9508071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/08071  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/12588  
 ; FILING DATE: 23 DEC 1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/998,003  
 ; FILING DATE: 29 DEC 1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta B.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 32149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 102:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4650 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 495..4103  
 ; PCT-US95-08071-102  
 Alignment Scores:  
 Pred. No.: 3,25e-46 Length: 4650  
 Score: 881.50 Matches: 343

Percent Similarity: 38.23% Conservative: 146  
 Best Local Similarity: 26.82% Mismatches: 452  
 Query Match: 5.67% Indels: 339  
 DB: 5 Gaps: 46  
 US-09-916-849a-3 (1-2923) x PCT-US95-08071-102 (1-4650)  
 QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
 Db 513 AGCCCAAGGCGCTGGGGGGAAGCGCTACGCTGCTCCATGCTGTAGACGCTGCTC 572  
 QY 23 LeuLeuLeuProProProLeuLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42  
 Db 573 CTGCTGGCTCCATCCCAAG-----CCAGCCACTCGGTAGTGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrGlyLeuThrGlyHis 62  
 Db 614 CAAGGTCGCGAGAAACAGCCACCAACCTCT---CATTTG-----GAGCTTGC 661  
 QY 63 SerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGlyLeuThrGlyHis 82  
 Db 662 AGCCCACTATGTTT-----TCCAGATGTGGGCACTGTACAGCTAGAGTGGTGC 715  
 QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92  
 Db 716 CCCGTAACCTTCGCTGATGAGCAAGACAGTGAATTTTCAACCAAGACCTC----- 769  
 QY 93 TrpCysProGlnSerGlnAlaHisIlePro-----LeuProProAlaProGlnGlyCys 110  
 Db 770 -----CATGACCGTGAAGGGGCTCGGTGAATGCAGAACACAGCT 808  
 QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleSerProGlnGlyIleValThr 130  
 Db 809 CCTGTGTATCCCTGCATCCCTGAGAGTTGAGT----- 841  
 QY 131 LeuProGlnGlnHisProCysLeuValAlaProArgLeuArgCysGlnSerCysIleLeu 150  
 Db 842 -----ATCATCACAGACCTCTCTGACAGATGC----- 868  
 QY 151 AlaglnAlaProGlyLeuArgAlaGlyIleuArgSerProGlnGlnSerLeuGlyArg 170  
 Db 869 ---GAGCCCGCGGCTGAGAGGGCCAGATGAGT----- 901  
 QY 171 ArgIlyArg-AsnValaenthrralaProGlnPheGlnProProserGlyGlnAlaThrVa 190  
 Db 902 ACAAGACATCAATAGACACAC-----CCCACTTGGCTCACCAGATCAGCATCTGGCAT 958  
 QY 190 lProGlnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206  
 Db 959 CCTTGAAGAACCAACATCGGCTCACTCTCCCATCCGCGTGGCTTCA----- 1007  
 QY 206 eAspProAspGlnGlyGlnAlaGlyArgLeuGlnThrThrMetAspAlaLeuPheAsp 226  
 Db 1008 -GACCTGATGCTGTCCCAAGGAGTGGCATCTTATAGCTGAGCGTGGCAGAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlnG 246  
 Db 1067 GGAGGAGAAGCA-----CCACAGCTCATTTGATGGGCA 1102  
 QY 246 uLeuAspArgGlnThrIlySerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
 Db 1103 CCTGACCGGTAGCGCTGAGACTCTTATGACCTCACTCAAGGTGACAGTGGCGGAG 1162  
 QY 266 tProArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAspAspHisAs 286  
 Db 1163 CCCCCAAGCGCCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1222  
 QY 286 pProValPheGlnGlnGlnGlnIlyArgGlnSerLeuArgGlnGlnGlnGlnGlnGlnG 306  
 Db 1223 CCCCAGTTTGAAGGCGCTCTTATGAGCGCACTATCTAGATAGATAGATAGATAGATAG 1282  
 QY 306 rGlnValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeu 326

Db 1283 CTCGTCATCCAGGGAAGCCAAATGACTCAGACCAAGGTGCAATGCAAGAAATGCAATA 1342  
 Qy 326 rArgLeuGluGluGlySerGlySerProSerGluVal-----PheGluIleAs 343  
 Db 1343 CACATTCACAG-----GCGCCGAAGTTGTAGGCGCTCTTCACAGTGA 1390  
 Qy 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTy 363  
 Db 1391 CAGAAACACTGACCTTATCACTGTTCAGAGCCCGGTGACCTGATGAGACCTTACACCCCT 1450  
 Qy 363 rGluLeuThrValGluIleAspArgGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCAGCTGTCTGTGAAGACCGAGGCCACCA-----CCCAAGTGTCCCGTGC 1504  
 Qy 383 aAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPhe----- 399  
 Db 1505 CCAGGTGTGTGACCGTGAAGACATGAATGACATGCCCCCACTTGATGATCCGGGG 1564  
 Qy 400 -----SerGluTyArgTyValValGluValArgGluAspValThrProG 415  
 Db 1565 CATAGGCTAGTACTCATCAAGATGGATGGATGCTCAAGATGTGGCAGAGA 1624  
 Qy 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValIle 435  
 Db 1625 GACAGCTGTGGCCCTGTGCAAGTGTCTGACCGAGATGAGGAGAGAAATGCAGCTGTC-- 1682  
 Qy 435 sTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyIleuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGGAGGTGATGTGCTTCACGTGCGCAGGCAAGCATGTGAGACAGGAG 1741  
 Qy 455 a-----LeuAspValValSerProLeuAspTyArgIleThrThrTy 468  
 Db 1742 TGACAGCAAGAAAGATATTCCTGCAACATCAACCCCGCTAGCTACGAAAGTCA 1801  
 Qy 468 sGluTyThrIleuArgValArgAlaGlnAspGlyGlyArgProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCATTTGAGTGTGGCTGAGCTGTGGCAACCCCACTCCAGACATTA 1861  
 Qy 488 rGlyLeuValThrValGluIleuAspIleAspAspAsnAlaProIlePheValSerTh 508  
 Db 1862 CTCCTC--AAGGTGAGGTGTGTGACGTCAATACACGACCTGTCTTCACTCAGAG 1918  
 Qy 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyIleuValLeuHsValG 528  
 Db 1919 TGTCACTAGAGTCGCTCCCGAAGAACAAACAGCTGTGAAGTGTGCGATGATAC 1978  
 Qy 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyArgLeu----- 544  
 Db 1979 TGCCAGTATGTGCTGTGCTCTTAATGTGAGCTGTGTTAATCTCTGAGAGCTGAGCC 2038  
 Qy 545 -AlaGlyValGlyHsAspPheProPheThrIleAsnAsnGlyThrGlyTPIIleSerVa 564  
 Db 2039 GCGTCTAAGGCG-----CTCTCAACATCTCAACCCGAGACTGAGATGATCAGGT 2089  
 Qy 564 lAlaIleGluLeuAspArgGluGluValAspPheTySerPheGlyValGluIleArgAs 584  
 Db 2090 GAAGACATCTCGATTCGGGAACAGCGGAGAGCATATGATTGAAGGTGGTGGCAGCTGA 2149  
 Qy 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGAGTCTTACCTTCAGGCAAGCCACTGCTCTTGTCAATGTGCTGAGCTGCA 2209  
 Qy 604 nAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TACATGATGCCAAA----- 2225  
 Qy 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 Qy 644 rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225

Qy 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyTyIleGluArgGlnTyValLeuAl 684  
 Db 2225 ----- 2225  
 Qy 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704  
 Db 2225 ----- 2225  
 Qy 704 pAlaAsnThrPheArgProValPheGlnSerSerHisTyThrValAsnValAsnGluAs 724  
 Db 2226 -----TTATGTGATGTGCTCAACTTTCAGTGTGAGAA 2263  
 Qy 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCCAGCACTAGTTCAGTGGCAGTGGCAGTGTGATGATGATGAGACCAAGGGGAGAA 2323  
 Qy 744 nAlaArgIleThrTyPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 Db 2324 TGCCAGGTGACGCTCTCAGTGAGAGGACAAAGGTGACTTGTATTCAGAAATGGCAC 2383  
 Qy 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyArgIleAspGlnValSerTyThrIleuAl 784  
 Db 2384 AGGCACCATCTTATCCAGCTGAGCTTGTATTCAGAGCAACAAGCACTTACCTTCCA 2443  
 Qy 784 aIleThrAlaArgAspAsnGlyIleProGlnTySerAspThrThrTyIleuGluIleLe 804  
 Db 2444 GCTGAAGGCAAGTGAATGTGGCTCCCACTCCCTGACCTTACGTTGCTGATCAACATCA 2503  
 Qy 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyGlnGlySerVa 824  
 Db 2504 TGTGTGAGCAAGAAATGACAAACCACTTATATCACTGCCCTTCT--AACACCTTC 2557  
 Qy 824 lTyArgIleAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCACAGCTCTCAACCCCAAGACAGCTCTTGTGAGACGCTGACCGAGGTGGAGCGCA 2617  
 Qy 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyGlyAspAspG 860  
 Db 2618 GACATTGACTGTGTGTCAATGCCAGCTGATCTACAGCATTGAGGTGGCAACCT-- 2675  
 Qy 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrIleuArgLeuAs 880  
 Db 2676 -TATGACTCTTCAGATGTGGTCAATTCAGGTGCGCATCACTCGAGAGAGAAATGA 2734  
 Qy 880 pArgGluAsnValAlaGlnTyThrValLeuArgAlaTyAlaValAspTyGlyMetProP 900  
 Db 2735 GCGGCGCCACAGTGGCTTACACGCGCTGTGTGAAGTCAGTGACCGCGCAAGCCCTC 2794  
 Qy 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspProVa 920  
 Db 2795 ACGGTATGGCAAGCGTGTGCTCATCTTATGTCAATGACATCTGCGCAACCGACGCT 2854  
 Qy 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937  
 Db 2855 GCTGAGAACCTCTCGTGGCCACAGCTTGAC-----ACGCGCGTGA 2896  
 Qy 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955  
 Db 2897 TATGACATTT-----GCTGGGAGATCCAGAAATGAGCGCTCCAGAGCGCTGG 2944  
 Qy 956 -----IleMetTyGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePh 974  
 Db 2945 CAACATTCCTTGT 2996  
 Qy 974 eSerGlyGluLeuThrAlaLeuVal----- 982  
 Db 2997 -----GCCCTGGCGGTCTTGTGCGCTACTGACAGCAAGCGGAGGCCAAAAGTGTTA 3049  
 Qy 983 -----AspLeuAspTyArgIleAspArgProGluTyValLe 994  
 Db 3050 CCAGCTGTGAAGAGAGACCAAGACCTG--TATGCCCCCAAGCCAGTGGCAAGGC 3106

QY 994 UValIIeGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014  
 DB 3107 CTCGAAGGAGAAACAAAGCAAGGAGCAAGCAAGTCCCAAGCCCGTGAAGCCAGT 3166  
 QY 1014 UAspArgAsnAspAsnProProValIleuGlyAsnAspGluIleuPheAsnAsnTYrVa 1034  
 DB 3167 GGAAGACAGAGATGAGCGCGGCTGACAGAGTCCCTCAAGTTC----- 3209  
 QY 1034 LThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054  
 DB 3210 ---AACCTGATGAGCATGCGCCCTGGGAGACAGT---CCCCGATC---CACCGGCC 3256  
 QY 1054 cAspIleSerAspSerIleuThrTYrSerPheGluArgGlyAsnGluLeuSerLeuValLe 1074  
 DB 3257 C----- 3257  
 QY 1074 ULeuAsnAlaSerThrArgIleuLeuValLeuSerArgAlaLeuAspAsnAsnArgProLe 1094  
 DB 3258 -CTCACTACACCACAGGACAGCCCTGACCTGGCCGCGCACTATGCTTACCTCCACT 3316  
 QY 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111  
 DB 3317 GCCTTCATCCAGCTGACGCGCCAGTCACTCCTCAGCTCCAGAGAGACCAAGGTGTACA 3376  
 QY 1111 agln-CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131  
 DB 3377 GGACCTGCACCTGCAC-----ACACATTCG----- 3402  
 QY 1131 euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeuG 1147  
 DB 3403 -----TGGGCACTGGGAGACACACGCTCCAGGCTCTAGACAGT----- 3441  
 QY 1147 LyleuPheIIeGlnAlaValAlaIleThrLeuAlaThrProProAspHisValValAlP 1167  
 DB 3442 -----ACTCCGACTACAGCTACCGCAACCGCCCA----- 3474  
 QY 1167 heAsnVal-GlnAspArgThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186  
 DB 3475 --AAATACCCCAAGCAGCATTACTCACTCCGCGGCTCACTTC----- 3517  
 QY 1187 ValGlyGlnProGlyProGlyGlyGlyProPheLeuProSerGlu 1203  
 DB 3518 ---GGCCACCAAGCCAGCGCCAGAGCTGCAGAGACC---ATCCAGACACAG 3562  
 RESULT 12  
 US-07-998-003A-94  
 ; Sequence 94, Application US/07998003A  
 ; Patent No. 5643781  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 107  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; STREET: 20 South Clark Street  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/998,003A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 5643781and, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 30903

; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/346-5750  
 ; TELEFAX: 312/984-9740  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 94:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4104 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 495..3572  
 ; US-07-998-003A-94  
 Alignment Scores:  
 Pred. No.: 1,81e-45 Length: 4104  
 Score: 868.50 Matches: 287  
 Percent Similarity: 40.59% Conservative: 123  
 Best Local Similarity: 28.42% Mismatches: 362  
 Query Match: 5.59% Indels: 239  
 DB: 1 Gaps: 31  
 US-09-916-849a-3 (1-2923) x US-07-998-003A-94 (1-4104)  
 QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
 DB 513 AGCCCAAGGCCCTGGGGGCAAGGCTACGCTGCCCTCCATGCTGTACAGACTGCTCTC 572  
 QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42  
 DB 573 CTGCTGCTTCATCCCAAG-----CCAGGCCACTGGGTAGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuGlySProSerSerAla 62  
 DB 614 CAAGTGCCGAGAAACAGCCACCAACCTT---CATTTG-----GAGCTCGC 661  
 QY 63 SerAsnLeuThrLeuTYrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82  
 DB 662 AGCGACATATGTTT-----TCCAGATGTGGGCACTGTACAACTAGAGGTGGTGC 715  
 QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92  
 DB 716 CCCGTACTTCGGGTGATGACAGACAGGTGCATTTTCAACCAAGACCTC----- 769  
 QY 93 TrpCysProGluSerGlnAlaHisIlePro-----LeuProProAlaProGluGlyCys 110  
 DB 770 -----CATGACCGTGAAGGGGCTCGGAATGCCAGAACCAAGCT 808  
 QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleLeuSerProGlnGlySLeuThr 130  
 DB 809 CCTGTGATTCCTGTGATCTGTGAGTTTGAAGT----- 841  
 QY 131 LeuProGluGlnHisProCysLeuValProArgLeuArgCysGlnSerCysSLeu 150  
 DB 842 -----ATCTATCACAGACCTCGTGCAGAAATGC----- 868  
 QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyArg 170  
 DB 869 ---GAGCCCCCGGCTGTAAGGGCCAGATAGAAGT----- 901  
 QY 171 ArgLeuArg-AsnValAsnThrAlaProGlnPheGlnProProSerTYrGlnAlaThrVa 190  
 DB 902 ACAAGACATCAATGACACACA---CCCAATTGCGCTCACAGGTATCATCTGGCCAT 958  
 QY 190 LProGluGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206  
 DB 959 CCTGAGAACACCAACATCGGCTCACTTCCCATCCCGCTGGCTTCA----- 1007  
 QY 206 eaAPProAspGlnGlyGluAlaGlyArgLeuGluTYrThrMetAspAlaLeuPheAspSe 226  
 DB 1008 -GACCGTATGCTGTGCCAAGGTGTGCATCTATGAGCTGCAGGTGCAGAGACCA 1066

QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 DB 1067 GGAGGAGAACGA-----CCACAGCTCATTTGGATGGGCAA 1102  
 QY 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
 DB 1103 CTGTGACCGGTGAGCGCTGGAGCTCTTATGACCTCATCAAGGTGACGATGGCGGACG 1162  
 QY 266 tProArgSerSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAspHisAs 286  
 DB 1163 CCCCCACGGGCGACGAGTGCCTGTGCTGCTGACCGGTGCTTGAACAATGACAAAGCGC 1222  
 QY 286 pProValPheGluGlnGlnGluThrLysGluSerLeuAspGluAsnLeuGluAlaGlyTy 306  
 DB 1223 CCCCAGATTGAGGGCCCTCTTGAAGCCGCACTTCTGAGAAATGAGCCCATAGGCCA 1282  
 QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTy 326  
 DB 1283 CTGCGTATCCAGGTGAAAGGCCAATGACTCAGACCAAGTGCCAATGCAAGAAATCGAATA 1342  
 QY 326 rArgLeuLeuGluLysSerGlyLysSerProSerGluVal-----PheGluIleAs 343  
 DB 1343 CACATTCCACAG-----GGCCCGAAGTTGTAGGCGCTTCTTCGACTGGA 1390  
 QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTy 363  
 DB 1391 CAGGAACACTGGAATTATCACTGTTCAAGGCGCCGCTGAGCCGTAGAGACCTTAACACACCT 1450  
 QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 DB 1451 GCGCTTCTCAGTGTCTGTGTAAGACCGAGGCCACCA-----CCCAAGAGTGCCTCGTGC 1504  
 QY 383 aAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPhe----- 399  
 DB 1505 CCAAGTGTGTGACCGTAAGACATGAAATGAAATGACCCCGCCACCTTGAATCCGGGG 1564  
 QY 400 -----SerGluValArgTyValGluValArgGluAspValThrProG 415  
 DB 1565 CATAGGGCTAGTACTCATGAATGGAGTGGATGCTTAACATCTCAGAGAGATGGGACAGGA 1624  
 QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaValIleHis 435  
 DB 1625 GACACGTGTGGCCCTGTGTGACGTGTGTGACCGAAGTGAAGGAATGACGTGTC-- 1682  
 QY 435 sTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyLeuAspAlaGlnThrGlyAl 455  
 DB 1683 -ACCTGTGTGTGACAGGTGATGTGCTTCCAGCTGCGCAAGGCCAGTGAACAGGCGAG 1741  
 QY 455 a-----LeuAspValIleSerProLeuAspTyGluThrThrTy 468  
 DB 1742 TGACAGCAAGAAAGATTTCTGTGACATACCAACCCCGCTAAGCTACGAGAAAGTCA 1801  
 QY 468 sGluTyThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 DB 1802 AGACTTACACCTTGAATTTGGCTGTGACTGTGCAACCCCGCACTTCCACACACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508  
 DB 1862 CTCCTC---AAGGTGACAGTGTGTGACGTCAATACACGCACTGTCTTCACTAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyTyLeuValLeuHisValG 528  
 DB 1919 TGTCACTAGAGTGCCTTCCCGAAGAAACAAACAGCCTGTGAAAGTATGTCTGATAC 1978  
 QY 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyTyArgLeu----- 544  
 DB 1979 TGCAGTATGTCTGCTGTAACTGAGCTGAGCTGTATCTCTGAGAGCTGAGGCC 2038  
 QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyTyrIleSerVa 564  
 DB 2039 GGCCTCTAAGGCG-----CTCTTACACATCTCACCCGAGATGAGATGCAGGT 2089

QY 564 lAlaAlaGluLeuAspArgGluGluValAspPheTySerPheGlyValGluAlaArgAs 584  
 DB 2090 GAAGACATCTCGATCGGAAACAGCGGAGACTATGATGTTGAAGTGTGGCAGCTGA 2149  
 QY 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 DB 2150 CCGGGCAAGTCTTAGCTTCAGGGGACAGCACTGTCTTGTCAATGTCTGAGTGCAG 2209  
 QY 604 nAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624  
 DB 2210 TGACATGACCCCAA----- 2225  
 QY 624 aValGlyThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 DB 2225 ----- 2225  
 QY 644 rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 DB 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyTyLysGluArgGlyIntyValLeuAl 684  
 DB 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValAlaAsnValThrAs 704  
 DB 2225 ----- 2225  
 QY 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyThrValAsnValAsnGluAs 724  
 DB 2226 -----TTATGTGTGAGTGTCTCAACTTCTCAGTGTGAGAA 2263  
 QY 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 DB 2264 CATGCCAGCATGAGTCCAGCTGGGCATGTGACTGTCATTGATGAGACAGAGGGGAGAA 2323  
 QY 744 nAlaArgGlyeThrTyPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 DB 2324 TGCCGAGGTGACGCTCTCAGTGAAGCAGCAACAGGTGACTTTGTATCCAGAAATGGCAC 2383  
 QY 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyGluAspGlnValSerTyThrThrLeuAl 784  
 DB 2384 AGGACACATCTATCCAGCTGAGCTTGTATGAGAGCAACAAAGCACTTACCTTCCA 2443  
 QY 784 aIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyLeuGluIleLe 804  
 DB 2444 GCTGAAGGAGTGGATGGTGGGTGCCACCTGCTCAGCTTACGTTAGTGTACACATCAA 2503  
 QY 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyGlnIlySerVa 824  
 DB 2504 TGTGCTGAGACGAAATGACAAACGCACTTATCACTGCCCTTCT---AACACCTC 2557  
 QY 824 lTyGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 DB 2558 TCAACAGCTCTACCCCGACAGCAAGCTCTTGTGTGAGACGCTACAGCCAGGTGCAAGCCA 2617  
 QY 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyGlyAspAspG 860  
 DB 2618 GACTTGAATCTGGTGTCAATGCGAGCTGATCTACAGATTGACAGTGGCAACCT-- 2675  
 QY 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAs 880  
 DB 2676 -TAGACTCTTCAATGTGGTCAATTCAGGTGCATCACTGAGCACTGAGAAAGAAATGA 2734  
 QY 880 pArgGluAsnValAlaGlnTyThrValLeuArgAlaTyAlaValAspLysGlyMetProP 900  
 DB 2735 CCGCGCGCACAGGGGTACACCGCTGTGTGTGAAGGTGACAGACCGCGGCAAGCCCC 2794  
 QY 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVa 920  
 DB 2795 ACGCTATGACAGAGCTTGTCTATGTCAATGACACTGTGCCCAACGACGCT 2854  
 QY 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937





Db 1505 CCAAGGTGGTGTGACCGTGAAGACATGATGACATGCCCCCACTTGAGATCCGGGG 1564  
 Qy 400 -----sergluysa-rgyrvaiValgluValargluaspValThrProgl 415  
 Db 1565 CATAGGGCTACTGACTCATCAAGATGGATGGCTTAACATCTCAGAGATGGTGGACAGGA 1624  
 Qy 415 yalaProvalleuargValThrAlaSerAspArgAspIysSeranaValValHt 435  
 Db 1625 GACAGCTGTGGCCCTGGGTGACAGTGTCTGACCGAGATGAGAGAGATGACGCTGC-- 1682  
 Qy 435 sTySerilleMetSerIysAnaIaarggluInpHeterIyeuaspIaagInThrIgl 455  
 Db 1683 -ACCTGTGTGGTGGAGTGTGCTGCTTCCAGCTGGCCAGGCGAGTGAAGAGGCTGAG 1741  
 Qy 455 a-----LeuaspValIserProleuaspIyrgluThrIly 468  
 Db 1742 TGACAGCAAGAAAGATTTCTGCGACATCAACCCCGCTTAACCTACAGAAAGCTCA 1801  
 Qy 468 ggluTyThrIleuargValArgAlaIaspIyglIyArgProProleuSerAnaIse 488  
 Db 1802 AGACTACACCACTGAGATTGTGCTGTGACTGTGCAACCCCGCTTCCAGCACTAA 1861  
 Qy 488 rgluIleuValThrValgluValleuaspIleaspAspAnaIaProIlePheValIserTh 508  
 Db 1862 CTCCCTC--AAGGTGAGGTGTGTGAGCTGATGACAGCACTGTGCTTCACTGACAG 1918  
 Qy 508 rProPhegluAlaThrValleuIleuIserValProleuGlIyIreueValleuHtIseVal 528  
 Db 1919 TGTCACGTGAGGTGCCCTTCCGGGAAACAAACAACCGCTGTGAATGCTGTGAGATCAC 1978  
 Qy 528 naIaIleaspAlaaspAlaagIyaspAnaIaargleuIyIyargleu----- 544  
 Db 1979 TGCCAGTGTGCTGCTGCTGCTTAATGTGAGCTGATTACTCTCTGGAGCCTGAGCC 2038  
 Qy 545 -AlaIglValIglIyIseAspPheProPheThrIleaspAnaIyIyHrdIyTPTIleSera 564  
 Db 2039 GCGTGTCAAGGAC-----CTCTTCAACATCTCAACCGAGCTGAGATGCAAGT 2089  
 Qy 564 lAlaAlaIleuIleuaspArgIleuIleuValaspPheTySerPhegluValgluAlaArgAs 584  
 Db 2090 GAACACATCTCTGATGGGGAACAGCGGAGAGAGCTGATGAGTGAAGGTGGTGGACGTGA 2149  
 Qy 584 pHiGlyThrProAlaIleuThrAlaSerAlaSerValIserValThrValleuaspValas 604  
 Db 2150 CCGGGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2209  
 Qy 604 naSPAspAnaProThrPheThrInProgluTyThrValargleuanaIaspAlaI 624  
 Db 2210 TGACATGATGACCCAAA----- 2225  
 Qy 624 aValIglIyIserValIleuThrValIserAlaValaIaspArgAspAlaIleIserValIleTh 644  
 Db 2225 ----- 2225  
 Qy 644 rTyrgluInIleThIsergluIleuThrArgAspArgPheSerIleThIsergluInsergluIgl 664  
 Db 2225 ----- 2225  
 Qy 664 yglIleuValIserIleuAlaIleuProleuaspTyIyIleuIleuIyrgluInTyValleuAl 684  
 Db 2225 ----- 2225  
 Qy 684 aValIleuValIserAspIyIleuThrArgIleuAspThrAlaIleuValaIleuValaIleuThrAs 704  
 Db 2225 ----- 2225  
 Qy 704 palAaenThrIleArgProvalPhegluIleuSerIleTyThrValaIleuValaIleuIleu 724  
 Db 2226 -----TTATGCTGAGTGGCTACCACTTCTCAAGATGAGAGAA 2263  
 Qy 724 pArgProAlaIyIleuThrValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 744  
 Db 2264 CATGCAAGCACTGAGTCAAGTGGGATGTGATCTGTATGTGAGAGCAAGGGGAGAA 2323

Qy 744 naIaArgIleuThrTyPheMetgluIleuaspSerIleProgluInpHearglIaaspAlaaspTh 764  
 Db 2324 TGCCAGAGTGCAGCTCTCAGTGGAGAGAGCAACAGGTGACTTTGTATTCAGAAATGGCAC 2383  
 Qy 764 rgluIleuValIleuThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 784  
 Db 2384 AGGCACATCTTATCCAGCTGAGCTTGTATGAGAGCAACAAAGCACTTACCTTCCA 2443  
 Qy 784 aIleThrIleuArgAspAnaIyIleProgluIleuIleuIleuIleuIleuIleuIleuIleu 804  
 Db 2444 GCTGAAGGAGTGGATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2503  
 Qy 804 uValaIleuValaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 824  
 Db 2504 TGTCGTGACAGAGATGACAGCACTTATATCACTGCCCCCTTCT-----AACACCTC 2557  
 Qy 824 lTyrgluIleuValProPheThr-----SerValleuIleuIleuIleuIleuIleuIleu 840  
 Db 2558 TCACAGCTGCTGACCCCGAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2617  
 Qy 840 rAspArgAspSergluIleuanaIyIleValPheTyThrPhegluIyglIyAspAspGl 860  
 Db 2618 GACCTTGTGCTGT 2675  
 Qy 860 yAspIyAspPheIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 880  
 Db 2676 -TATGACCTTCCAGATTGGGTGCTCATTCAGGTGCCATCACTTGAAGAGAGATTTGA 2734  
 Qy 880 pArgIleuIleuValaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 900  
 Db 2735 GCGGCCCAACATGAGGCTACACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2794  
 Qy 900 oAlaIyGThrPrometgluIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 920  
 Db 2795 ACCCTATGACACAGCTTGT 2854  
 Qy 920 lPhegluIn-----AspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 937  
 Db 2855 GCTGAGAGACCTTCTGAGGACACAGCCTGTGAC-----ACGCGCTGGA 2896  
 Qy 937 yIleuAlaValaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 955  
 Db 2897 TATGTACAT-----GCTGGAGATTCAGATATGACGCTCCAGAGCAGCTGG 2944  
 Qy 956 -----IleMetTyrgluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 963  
 Db 2945 CAACATTTCTTGT 2972  
 RESULT 14  
 US-08-453-695A-94  
 ; Sequence 94, Application US/08453695A  
 ; Patent No. 5708143  
 GENERAL INFORMATION:  
 APPLICANT: Suzuki, Shintaro  
 TITLE OF INVENTION: Proteocadherin Materials and Methods  
 NUMBER OF SEQUENCES: 115  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ADDRESSEE: Borun  
 STREET: 233 South Wacker, 6300 Sears Tower  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Releasee #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/453,695A  
 FILING DATE:

CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5708143and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32658  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 94:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4104 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 495..3572  
 US-08-453-695A-94

Alignment Scores:  
 Score: 1.81e-45 Length: 4104  
 Percent Similarity: 868.50 Matches: 287  
 Best Local Similarity: 40.59% Conservative: 123  
 Query Match: 5.59% Mismatches: 362  
 Indels: 239  
 Gaps: 31

US-09-916-849A-3 (1-2923) x US-08-453-695A-94 (1-4104)

QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
 Db 513 AGCCAGAGCCCTCGGGGAGCAAGGCTACTGCTCCCTCCAGCTAGACCTGCTC 572  
 QY 23 LeuLeuLeuProProProLeuLeuGlyAAspGlnValGlyProCyAAspSerLeuGlySer 42  
 Db 573 CTGCTGCTCTCATCCCAAGG-----CCAGCCCACTGGGTAGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCyAProSerSerAla 62  
 Db 614 CAGAGTCCCGAGAGAACAGCCACCAACCTT---CATTCG-----GAGCTCCG 661  
 QY 63 SerAsnLeuTrpLeuTyTrpSerArgCyAAspAspAlaGlyThrGluLeuThrGlyHis 82  
 Db 662 AGCCGATATGATGTTT-----TCCAGATGTGGGCACTGTACCAAGCTAGAGTGGGTGC 715  
 QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92  
 Db 716 CCCGTAACCTTCGCGTGGATGGCAAGAGTGCATTTTCAACCAAGGAGCTC----- 769  
 QY 93 TrpCyAProGluSerGluAlaHisAlaPro-----LeuProProAlaProGluGlyCys 110  
 Db 770 -----CATCGACCGGTAGAGGGCTCCGTGAATGCCAGAACCAAGCT 808  
 QY 111 ProTrpSerCyAAspLeuLeuGlyTleGlyGlyHisLeuSerProGluGlyLeuThr 130  
 Db 809 CCTGTGGATATCCCTGCATCTCTGAGGTTTGAAGT----- 841  
 QY 131 LeuProGluGluHisProCyAAspLeuValAlaProArgLeuArgCyAAspSerCyAAspLeu 150  
 Db 842 -----ATCTATCAGACACTGCTGACAGAAATGC----- 868  
 QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyGlyArg 170  
 Db 869 ---GAGCCCGCGGTGCTAGAGGCGCCAGATGAAGT----- 901  
 QY 171 ArgLeuArgAspAsnValAlaSerThrAlaProGluPheGlnProProSerTyGlnAlaThrVal 190  
 Db 902 ACAAGACATCAATGACACACACA-----CCCAACTTCGCTCCACATCATCACTGGCCAT 958  
 QY 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaVal 206  
 Db 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaVal 206

Db 959 CCTGAGAACACCAACATCGGCTCACTCTTCCCATCCCGCTGCTCA----- 1007  
 QY 206 eAAspProAspGluGlyGluAlaGlyArgLeuGlyTrpThrMetAspAlaLeuPheAsp 226  
 Db 1008 -GACCGTGAATCTGGTCCCAACGGTGTGGCATCTTATGACTCCAGGTGCAAGAGCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 Db 1067 GGAGAGAGAACCA-----CCACACTCATTTGTGATGGCAA 1102  
 QY 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGly 266  
 Db 1103 CCGTGAACCGGTAGCGGTGGAGCTCTATGACTCACCACATCAAGGTGAGATGGCGCAG 1162  
 QY 266 tProArgAspSerAlaLeuAlaThrLeuThrLleuValThrAspThrAspHisAsp 286  
 Db 1163 CCCCCAGCGGCAACAGATGCTCTGCTGGGTGTCACCGGTCTTGACCAACATACAAAGC 1222  
 QY 286 pProValPheGluGlnGlnGlyTrpLysSerLeuArgGluAsnLeuGlyValGly 306  
 Db 1223 CCCCAGTTTGAAGCGGCGCTCTATAGAGCCGAACCTATCTGAGAAATAGCCCAATAGGCCA 1282  
 QY 306 rGluValLeuThrValAlaArgAlaThrAspGlyAspAlaProProAsnAlaAsnLleu 326  
 Db 1283 CTGCTGCTACAGGTGAGAGGCAATGACTCAGACCAAGGTGCCAATGCAAAATTCAAAT 1342  
 QY 326 rArgLeuLeuGluGlySerGlyLysProSerGlyVal-----PheGluLys 343  
 Db 1343 CACATTCACACAG-----GGCCCGAAAGTTGAGAGCGTCTTCCGACTGGA 1390  
 QY 343 pProArgSerGlyValAlaArgThrArgGlyProValAspArgGluGluGlySer 363  
 Db 1391 CAGAGACACTGAGCTTATCATCTGTCAGAGGCCCGGTGAGACCGTGAAGACTTAAGACCCCT 1450  
 QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCTCAGTCTTGTCTAAGACCGAGGCAACCA-----CCCAAGATGCCCCGTGC 1504  
 QY 383 aAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPhe----- 399  
 Db 1505 CCAAGGTGTTGGACCTGTAAAGGACATGAATGCAATGCCCCCAACATTGAGATCCGGGG 1564  
 QY 400 -----SerGluAspArgTyValAlaGlnValArgGluAspValThrProG 415  
 Db 1565 CATAGGGCTAGTACTCATCAAGATGGAGTGGTAACTCATCAAGAGATGTGCAGAGGA 1624  
 QY 415 yAlaProValLeuAspValThrAlaSerAspArgAspArgGlySerAsnAlaValHis 435  
 Db 1625 GACAGCTGTGGCCCTGTGTCAAGGTGTCTGACCGAGATGAGGAGAAATGCAAGCTGTC-- 1682  
 QY 435 sTySerLleMetSerGlyAsnAlaArgGlyGlnPheTyTrpLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGGAGAGTGTATGCTCCATCAAGTCCGCCAGGCGCAAGTGAAGAGGCAAG 1741  
 QY 455 a-----LeuAspValAlaSerProLeuAspTyTrpGluThrThrLy 468  
 Db 1742 TGACAGCAAGAAAGATATTTCTGTGACACTACACCCCGGTAGACTAGAGAAAGTCAA 1801  
 QY 468 sGluTyTrpThrLeuAspValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCAATGATGATGTGGCTGTGACCTGTGCAACCCCACTCTCCAGACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspLleAsnAspAsnAlaProLlePheValSerTh 508  
 Db 1862 CTCCTCC---AAGTGTGAGGTGTGAGCTCATATGACAAAGCACTCTTCACTCAAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyTrpLeuValLeuHisValG 528  
 Db 1919 TGTCACTGAGTGTGCTTCCCGAAACAAACAAAGCGTGTGAGTGAAGTGAATGCTGAGATCAC 1978  
 QY 528 nAlaLleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTrpArgLeu----- 544  
 Db 1979 TGCAGATGATGCTGACTGTGGCTTAATGTGAGCTGTGTTACTCTCTGAGACCTGAGGCC 2038

```

QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyrPheSerVa 564
Db 2039 GGCTGCTAAGGCGC-----CTCTTCAACATCTTCAACCCGAGACTGAGAGATCCAGCT 2089
QY 564 1AlaAlaGlyLeuAspArgGlyGluValAlaAspPheTyrSerPheGlyValGluAlaArgAs 584
Db 2090 GAAGACATCTCTGGATCGGGAACCGCGGAGAGCATATGATGAAGGTGGGCGAGCTGA 2149
QY 584 PheGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604
Db 2150 CCGGGGCGAGCTCTGAGCTCCAGGCGACAGCATCTGTCCTTCAATGCTGAGCTGCA 2209
QY 604 nAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsnGluAspAlaAl 624
Db 2210 TGACATGACCCCA----- 2225
QY 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644
Db 2225 ----- 2225
QY 644 rTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGly 664
Db 2225 ----- 2225
QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGlnTyrValLeuAl 684
Db 2225 ----- 2225
QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704
Db 2225 ----- 2225
QY 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAs 724
Db 2226 -----TTTATGCTGAGTGGCTTCAACTTCTCAAGATGAGACAA 2263
QY 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744
Db 2264 CATGCGACAGCTGAGTGCAGTGGCATGTGATCTGTCAATGATGAGCAAGGGGAGACAA 2323
QY 744 nAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764
Db 2324 TGCCCAAGGTGACGCTCTCGATGAGACAGACAAAGGTGATCTTGTATTCAGAAATGGCAC 2383
QY 764 rGlyValAlaValThrThrGlnAlaGluLeuAspTyrGlyAspGlnValSerTyrThrLeuAl 784
Db 2384 AGGCACCATCTCATCTGACGCTGATGATGAGAGACAAAGACACTTACACTTCCA 2443
QY 784 aIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyrLeuGluIleLe 804
Db 2444 GCTGAAGGACAGATGATGGTGGCTCCACCTGCTCAGCTTACGTTGGTGGACCATCA 2503
QY 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerVa 824
Db 2504 TGTGCTGACGAGATGACAAAGCAAGCCCTATATCACTGCCCTCT-----AACACCTC 2557
QY 824 lTyrGluAspValProProPheThr-----serValLeuGlnIleSerAlaTh 840
Db 2558 TCACAAAGCTGACGCCCCCAACACAGCTTGTGTGATGACAGCTGACGAGTGGCGACCGCA 2617
QY 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db 2618 GGAATTTGACTCTGGGTGTCATGCGAGCTGATCAAGCATTTGACGAGTGGCAACCCCT-- 2675
QY 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAs 880
Db 2676 -TATGACTCTTCCAGATGGGTGCATTCAGATGCTGACCTGAGAGAGAGATTGA 2734
QY 880 pArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIlyGlyMetProPr 900
Db 2735 GGGGGGCCACAGGAGCTACACGCGCTGGTGTGAAGTCACTGACCGGGGCAAGCCGCC 2794

```

```

QY 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVa 920
Db 2795 ACCGTATGGACAGCCTTGCTCATCTTATGTCATGATGAGACTTGGCCAAAGCGAGCT 2854
QY 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937
Db 2855 GCTGAGAGACCTCTGCGGCGACAGCCTGAGC-----ACGCCGCTGGA 2896
QY 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955
Db 2897 TATTGACATT-----GCTGGGATTCAGAAATATGACCGCTCCAGAGAGCTGG 2944
QY 956 -----lMetTyrGlnIleValGluGly 963
Db 2945 CAACATCTCTTGTGTGTGTGCTGCTGT 2972

```

# RESULT 15

US-08-268-161A-94

Sequence 94, Application US/08268161A

Patent No. 5798224

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Procoadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/268,161A

FILING DATE: June 27, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Young J. Sun

REGISTRATION NUMBER: P-41,337

REFERENCE/DOCKET NUMBER: 27866/32149

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 94:

SEQUENCE CHARACTERISTICS:

LENGTH: 4104 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 495..3572

US-08-268-161A-94

Alignment Scores:

Pred. No.: 1,81e-45

Score: 868.50

Percent Similarity: 40.59%

Best Local Similarity: 28.42%

Query Match: 5.59%

DB: 1

US-09-916-849A-3 (1-2923) x US-08-268-161A-94 (1-4104)

QY 3 serProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 03:42:22, Search time 1948 Seconds  
(Without alignments)  
5527.340 Million cell updates/sec

Title: US-09-916-849a-3  
Perfect score: 15545  
Sequence: 1 MMSPATGVLPPLPPPLLL.....AGTVDESSGSEFLFFNLH 2923

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US09916849.rnmat.11022004.143818.21031/asp.query.fasta.1.3079  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=BIOSUM62  
-TRANS=human40.csi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USERS=US09916849 @CGN 1.1 1417 @rnmat.11022004.143818.21031  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBIOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCT05\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	15545	100.0	8772	9	US-09-788-711A-3	Sequence 3, App1
2	15545	100.0	10531	15	US-10-225-567A-523	Sequence 523, App
3	15518.5	99.8	8871	1	US-09-788-711A-1	Sequence 1, App1
4	12339	79.4	24370	12	US-10-292-798-931	Sequence 931, App
5	8781.5	56.5	9045	9	US-09-737-149-1	Sequence 1, App1
6	8781.5	56.5	11389	9	US-09-728-711-4	Sequence 4, App1
7	8781.5	56.5	11389	12	US-10-295-027-749	Sequence 749, App
8	8781.5	56.5	11389	12	US-10-240-145-20	Sequence 20, App1
9	8781.5	56.5	11389	13	US-10-241-220-51	Sequence 443, App1
10	8781.5	56.5	11389	15	US-10-225-567A-443	Sequence 443, App1
11	8736.5	56.2	9087	12	US-10-150-811-13	Sequence 13, App1
12	8736.5	56.2	9087	13	US-10-131-409-13	Sequence 13, App1
13	7836.5	50.4	11965	15	US-10-225-567A-655	Sequence 655, App
14	7766	50.0	11648	15	US-10-149-819-42	Sequence 42, App1
15	3958	25.5	2332	9	US-09-764-870-569	Sequence 569, App
16	3958	25.5	2332	15	US-10-125-540-569	Sequence 569, App
17	3888	24.6	2291	15	US-10-176-847-99	Sequence 99, App1
18	3787	24.4	2603	9	US-09-843-856-1	Sequence 1, App1
19	3734	24.0	4152	10	US-09-925-300-359	Sequence 359, App
20	3701	23.8	2695	12	US-10-264-237-636	Sequence 636, App
21	3526	22.7	2077	9	US-09-764-870-570	Sequence 570, App
22	3526	22.7	2077	15	US-10-125-540-570	Sequence 570, App
23	3318.5	21.3	3492	12	US-10-150-811-67	Sequence 67, App1
24	3318.5	21.3	3492	13	US-10-131-409-67	Sequence 67, App1
25	3317	21.3	3493	13	US-10-150-811-68	Sequence 68, App1
26	3317	21.3	3493	13	US-10-131-409-68	Sequence 68, App1
27	2070.5	13.3	11002	13	US-10-017-161-1095	Sequence 1095, App
28	1732.5	11.1	15603	13	US-10-114-153-63	Sequence 63, App1
29	1520.5	9.8	14756	12	US-10-295-027-1106	Sequence 1106, App
30	1520.5	9.8	14756	15	US-10-160-758-5	Sequence 5, App1
31	1520.5	9.8	14756	15	US-10-060-036-134	Sequence 134, App
32	1460	9.4	1376	13	US-10-029-386-22961	Sequence 22961, App
33	1451	9.3	925	9	US-09-764-870-70	Sequence 70, App1
34	1451	9.3	925	15	US-10-125-540-70	Sequence 70, App1
35	1443	9.3	14334	15	US-10-084-817-227	Sequence 227, App
36	1432	9.2	14749	13	US-10-210-120-72	Sequence 72, App1
37	1423.5	9.2	9294	13	US-10-174-677-112	Sequence 112, App1
38	1415.5	9.1	13700	12	US-10-085-198-21	Sequence 21, App1
39	1338.5	8.6	14536	13	US-09-970-944-3	Sequence 3, App1
40	1338.5	8.6	14536	15	US-10-160-758-6	Sequence 6, App1
41	1338.5	8.6	14536	15	US-10-160-758-7	Sequence 7, App1
42	1331	8.4	14930	12	US-10-062-674-1800	Sequence 1800, App
43	1306.5	8.4	11073	12	US-10-085-117-113	Sequence 113, App
44	1306.5	8.4	11073	15	US-10-160-758-2	Sequence 2, App1
45	1305	8.4	10065	12	US-10-085-117-110	Sequence 110, App

## ALIGNMENTS

RESULT 1  
US-09-788-711A-3  
Sequence 3, App1 Application US/09788711A  
Patient No. US20020058328A1  
GENERAL INFORMATION:  
APPLICANT: Tania Tamsin Testa  
TITLE OF INVENTION: NOVEL COMPOUNDS  
FILE REFERENCE: GP-30225  
CURRENT APPLICATION NUMBER: US/09/788, 711A  
CURRENT FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 0004196.2  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 8772  
TYPE: DNA  
ORGANISM: HOMO SAPIENS  
US-09-788-711A-3

Alignment Scores:  
Pred. No.: 0  
Score: 15545.00  
Percent Similarity: 100.00%  
Length: 8772  
Matches: 2923  
Conservative: 0

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-916-849A-3 (1-2923) x US-09-788-711A-3 (1-8772)

QY	1	LeIaAsrLeuProAlaThhGlnValProleuProThrProProProleuLeuLeu	20
Db	1	ATGGGAGACCGGACCAACGGAGCTCCCTCCCAAGCGCGCGCGCTCTGTGTG	60
QY	21	LeuLeuLeuLeuLeuProProProleuLeuGlyAapGlnValGlyProCyArGSerLeu	40
Db	61	TTGCTGCTGCTGCTCCCGCCCACTATTGGGAGACCAAGTGGGGCTGTGTTCTTG	120
QY	41	GlySerArgGlyYArgGlySerSerGlyValaCysAlaPrometGlyTyrPleuCyEProSer	60
Db	121	GGGTCCAGAGGAGCGAGGCTCTTCGGGGGCGCTGGCGCCCAATGGGCTGTGTCATCC	180
QY	61	SerAlaSerAemLeuTyrPleuTyrThrSerArgCysArgAapAlaGlyThrGlnLeuThr	80
Db	181	TCAGGTCGAACCTCTGGCTCTTACACAGCGGCTGACGAGGATGGGGGCACTGAGTGA	240
QY	81	GlyHisLeuValProHisHisAapGlyLeuArgValTyrCyEProGlnSerGlnAlaHis	100
Db	241	GGCCACCTGGTACCCCAACAGCATGCGCTGAGGGTTTGATGCCAAATCCGAGGCCAT	300
QY	101	IleProleuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyIleGly	120
Db	301	ATTCCCTTACCAACCAAGCTCTGTGAAGGTGGCCCTGAGACTGTGGCTCTGGGCA	360
QY	121	GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnHisProCysLeuValAla	140
Db	361	GGCCACCTTTCCCAACAGGCAAGTCAACCTGCCCGAGAGACACCTGTGCTTAAAGCT	420
QY	141	ProArgLeuArgCysGlnSerCysIleValAlaGlnAlaProGlyLeuArgAlaGlyGlu	160
Db	421	CCAGGCTCAAGTGCAGTCCGTGCAAGGCTGGCAAGGCCCGCGGCTCAGGGCAGGGGA	480
QY	161	ArgSerProGlnGlnSerLeuGlyGlyValArgArgValArgAemValAemThrAlaProGln	180
Db	481	AGGTACCAAGAAAGTCTCCCTGGGTGGGTGGGAAAGAAATGTAATAACAGCCCCCAG	540
QY	181	PheGlnProProSerTyrGlnAlaThrValProGlnAemGlnProAlaGlyThrProVal	200
Db	541	TTCCAGCCCCCAAGCTACAGGCGCACAGTGGCGGAGAACAGCGCAGAGGACCCCTGT	600
QY	201	AlaSerLeuArgAlaIleAapProAapGlnGlyGlnAlaGlyValGlnGlnGlyThrMet	220
Db	601	GCATCCCTGAGGGCCATGACCCCGACAGGGGTAGGAGGTGCACCTGAATGACCAATG	660
QY	221	AapAlaLeuPheAapSerArgSerAemGlnPhePheSerLeuAapProValThrGlyAla	240
Db	661	GATGCCCTCTTGAATAGCCGCTCCAACACAGTCTCTCCCTGGACCCAGTCACTGTGTCA	720
QY	241	ValThrThrAlaGlnGlnLeuAapArgGlnThrGlySerThrHisValPheArgValThr	260
Db	721	GTAAACACAGCCGAGAGAGCTGTGATGTGTAGAACCAAGAGACCCCAAGTCTTCAGGGTCA	780
QY	261	AlaGlnAapHisGlyMetProArgArgSerAlaAlaAlaThrLeuThrIleLeuValThr	280
Db	781	GGCGAGAGACAGCGCATGCCCGGAGAGATGCCCTGGCTACCACTCATCTTGGTTACT	840
QY	281	AapThrAapAapHisAapProValPheGlnGlnGlnGlyTyrGlyGlnSerLeuArgGlu	300
Db	841	GACACCAATGACCAATGACCTGTGTTCAGACGACGAGATGCACAAAGAGAGCTCAGGGAG	900
QY	301	AemLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAapGlyAapAlaProPro	320
Db	901	AACCTGGAGCTTGGCTATAGAGTGTCACTGTCAAGGCCAACGATGTGATGCCCCCTCC	960
QY	321	AemAlaAemIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe	340
Db	961	AATGCCAAATATCTGTACCGCTGTGGAGGGGTCTGGGGGGAGGCCCTCTGAAATCTTT	1020

[illegible]

QY 701 AsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 Db 2101 AAGTCAACCGACCGCCAAACCCATCGCTCTTTCAGAGCTCCCATATACAGTGAAT 2160  
 QY 722 ValAsnGlnAspArgProAlaGlyThrThrValValLeuLysSerLalaThrAspGlnAsp 740  
 Db 2161 GTTAAAGAGACCGCGCGGACCGACCGGTGTGTATGATGACCGCACGGATGAGGAC 2220  
 QY 741 ThrGlyValAsnAlaArgLysThrTyrPheMetGlnAspSerLysProGlnPheArgLys 760  
 Db 2221 ACAGGTGAGATGCCCCGATCACCTACTTCATGAGAGACAGCATCCCGCATCCGCATC 2280  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeuAspTyrGlnAspGlnValSer 780  
 Db 2281 GATGACGACACCGGGGCTGTCAACCCAGCTGAGGTGATCAAGACCAAGGTCT 2340  
 QY 781 TyrThrLeuAlaLeuThrAlaArgAspAsnGlyLeProGlnLysSerAspThrThrTyr 800  
 Db 2341 TACACCTCGCCATTAAGTCTCGGACCATGGCAATCCCAAGAGTCCACACACCTAC 2400  
 QY 801 LeuGlnLysLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db 2401 CTGGAGATCTCGGTGAACGACGTAATGACATGCCCTCAGTTCCTGGGAGCTCTAC 2460  
 QY 821 GlnGlySerValTyrGlnAspValProProPheThrSerValLeuGlnLysSerLalaThr 840  
 Db 2461 CAGGGACGTGTATAGATGATGCCACCTTCACTAGCGTCTGACGATCTCAGCCACT 2520  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspArgLys 860  
 Db 2521 GATCGATTCGTGACTTAATGACAGGGCTTCTTACCTCCAGAGGCGACGATGGA 2580  
 QY 861 AspGlyAspPheLysValGlnSerThrSerGlyLysLalaArgThrLeuArgArgLeuAsp 880  
 Db 2581 GACGGGACTTATTTGTGAGTCAAGTCAAGCATGTGTGAGAACGTTCAGAGGCTGAT 2640  
 QY 881 ArgGlnAsnValAlaGlnTyrValLeuArgLalaTyrAlaValAspLysGlyMetProPro 900  
 Db 2641 CAGAGAAAGGTGGCCAGATGTCTTGGCGGCAATGACGTGACAAAGGAGATGCCCA 2700  
 QY 901 AlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProProVal 920  
 Db 2701 GCCCGACACTTATGAGATGACATGCTGTGTGATGTGATATACATCCCTGTC 2760  
 QY 921 PheGlnGlnAspGlnPheAspValPheValGlnGlnAsnSerProLysLeuAlaVal 940  
 Db 2761 TTGAGACAGATGAGTTGATGTGTTGTGAGAGAGAACAGCCCATTTGGGCTAGCCG 2820  
 QY 941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnLysMetTyrGlnLys 960  
 Db 2821 GCCCGGCTCACAGCCCTGACCCCGATGAAGGACCAATGCCCATTTATGACAGATT 2880  
 QY 961 ValGlnGlyAsnLysProGlnValPheGlnLeuAspLysPheSerGlyValLeuThrAla 980  
 Db 2881 GTGGAGGGCAATCCCTGAGGTCTTCAGCTGACATCTTCTCCGGAGAGCTGACAGCC 2940  
 QY 981 LeuValAspLeuAspTyrGlnAspArgProGlnTyrValLeuValLysGlnAlaThrSer 1000  
 Db 2941 CTGGTGAAGTTCATACAGAGACCGGCGCTGAGTACGTCTGTGATTCACAGGACAGT 3000  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 3001 GCCCTCTGTGTGCGCGGTACAGTCCAGTCCGCTCTTGAACCGCATGACACCCA 3060  
 QY 1021 ProValLeuGlyAsnPheGlnLysLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040  
 Db 3061 CCAAGTCTGGGCAACTTTCATGATCTTTCACACATATGTCAACATGTCTCAGACAC 3120  
 QY 1041 PheProGlnGlyValAlaLysArgValProAlaHisAspProAspLysSerAspSerLeu 1060  
 Db 3121 TTCCTGGGGGTGCCATTGGCGAGTACCTGCCCATGACCTGTATCTCAATATGCTCG 3180  
 QY 1061 ThrTyrSerPheGlnArgLysValAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080

Db 3181 ACTTACAGCTTGGACGGGGAAATGAACTCACCTGCTCTGCTCAATGCTCCAGGGGT 3240  
 QY 1081 GlnLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaLysSerVal 1100  
 Db 3241 GAGCTGAAGCTAAGCGCGACCTGACACAAACCGGCTCTGTAGAGCTATGAGCGG 3300  
 QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrLysLe 1120  
 Db 3301 CTGGTGTACAGAGGGGTACAGCGGTACAGCGGCTGAGGCGCTGTGTGACATCATC 3360  
 QY 1121 ThrAspGlnMetLeuThrHisSerLysThrLeuArgLeuGlnAspMetSerProGlnArg 1140  
 Db 3361 ACCGATGAGATCTCACCAACAGATACAGCTGCGCCCTGAGAGAAATGTCAACCGAGCGC 3420  
 QY 1141 PheLeuSerProLeuGlnGlyLeuPheLysGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
 Db 3421 TTCTGTACACATCTCATAGGCTCTTCAATCCAGGCGGTGGCCGACGTTGGCACGCTCA 3480  
 QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyValHisLysLe 1180  
 Db 3481 CCGACACAGTGTGTCTTCAACGTACAGCGGACACCGAGCCCGGGGGCCACATC 3540  
 QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyTyrGlyValProProPheLeu 1200  
 Db 3541 CTCAACGTGAGGCTGTCCGTGGGCGAGCGCCAGGGCGCGGGGCGCGCCCTTCTCG 3600  
 QY 1201 ProSerGlnAspLeuGlnGlnArgLeuTyrLeuAsnArgSerLeuThrAlaLysSer 1220  
 Db 3601 CCTCTAGAGACTCAGAGAGCGCTTATATCCCAACCGAGCTGTGACGGCCATCTCG 3660  
 QY 1221 AlaGlnArgValLeuProPheAspAsnLysCysLeuArgGlnProCysGlnAsnTyr 1240  
 Db 3661 GCACAGCGGTGTCTCTCTTGAACGACAAATCTGCTCGGGAGACCTCTGGAGAACTAC 3720  
 QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheLysLysSerSer 1260  
 Db 3721 ATGCCCTGTGTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780  
 QY 1261 ValLeuPheArgProLysIleProValGlyValLeuArgCysArgCysProProGlyPhe 1280  
 Db 3781 GTGCTCTTCCGGCCATACACCCGTCGAGAGGCTGTGCTGTGCTGTGCTGTGCTGTG 3840  
 QY 1281 ThrGlyAspTyrCysGlnThrGlnValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3841 ACGGTGACTACTGTGAGAACGAGGTGACCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 3900  
 QY 1301 GlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 Db 3901 GGGCGCTCCGACGGCGGAGGGCGGCTACACCTGCTGTGTGTGTGTGTGTGTGTGTGT 3960  
 QY 1321 GlnHisCysGlnValSerLysArgSerGlyArgCysThrProGlyValCysLysAsnGly 1340  
 Db 3961 GAGCACTGTGAGTATGCTGTGCTGTGAGGCGCTGTGCTGTGCTGTGCTGTGCTGTG 4020  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPhe 1360  
 Db 4021 GGCACTGTGTCAACTGT 4080  
 QY 1361 GlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheLysThrPhe 1380  
 Db 4081 GAGAGGCTCTATGCTGACAGTACACAGCGCAGCTTCCCGCCATCTCTTATCACTCTT 4140  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGlnArg 1400  
 Db 4141 CCGGCGCTCGCGCACGCTTTCATCTTCACTTCACTTCCGCTCTGTGTGTGTGTGTGTGT 4200  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlnLysHisAspPheValAlaLeuGln 1420  
 Db 4201 GACGGGT 4260  
 QY 1421 ValLysGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440

Db 4261 GTATCCAGAGCAGATCCAGTCACTTCTCTGACGGGAGTCAACCAACGAGTGTCC 4320  
 Qy 1441 ProheValProGlyValSerAspGlyValIleThrIleThrValGlnIleuValTyrTyr 1460  
 Db 4321 CCATTCGGCCCGGAGAGTCAAGTATGGCCAGTGGCCATACGGTGCAGCTAAATACATAC 4380  
 Qy 1461 AsnIleProLeuLeuGlyValIleThrGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
 Db 4381 AATTAAGCAGCTGTTGGGTCAACAGAGGCTCCCAAGGGCCCATCAAGCAGAAAGTGGCT 4440  
 Qy 1481 ValValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 Db 4441 GTGGTGAACCTGT 4500  
 Qy 1501 AsnIleSerCysValAlaGlnGlyIleThrGlnGlyValSerValSerLeuAspLeuThr 1520  
 Db 4501 AACTACTCTGT 4560  
 Qy 1521 G1yProLeuLeuLeuGlyValValProAspLeuProGlySerPheProValArgMetArg 1540  
 Db 4561 GGGGCCCTGCTACTAGGGGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4620  
 Qy 1541 GlnPheValGlyCysMetArgAspLeuGlnValAspSerArgIleIleAspMetAlaAsp 1560  
 Db 4621 CAGTTCGTGGGT 4680  
 Qy 1561 PheIleAlaAspAsnGlyIleThrValProGlyCysProAlaValIleValAspValCysAspSer 1580  
 Db 4681 TTCAATTCGCAACATATGCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4740  
 Qy 1581 AsnThrCysValIleAsnGlyValIleThrCysValAlaAsnIleThrAspAlaPheSerCysGlyCys 1600  
 Db 4741 AACACTTCGCAACATATGCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
 Qy 1601 ProLeuGlyPheGlyValIleSerCysValAlaGlnIleMetAlaAspProGlnIlePheLeu 1620  
 Db 4801 CCCCTGGGCTTTGGGGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4860  
 Qy 1621 GlySerSerLeuValAlaIleThrIleGlyLeuSerLeuProIleSerGlyProThrTyrLeu 1640  
 Db 4861 GGCAGACGCTGT 4920  
 Qy 1641 SerLeuMetPheArgThrArgGlnIleAspGlyValIleLeuLeuGlnIleIleThrArgGly 1660  
 Db 4921 AGCCTCATGTTCGCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980  
 Qy 1661 ArgSerThrIleIleThrLeuGlnIleuArgGlnIleValIleValMetLeuSerValGlnGlyIle 1680  
 Db 4981 CGCAGCAGCAGTCACTTCAAGCTAGAGAGGAGCAGTGTGTGTGTGTGTGTGTGTGTGTGT 5040  
 Qy 1681 GlyLeuGlnIleAspSerLeuArgLeuGlnProGlyValArgAlaAspAspGlyValIleThrIle 1700  
 Db 5041 GGGGCTTCAGGCT 5100  
 Qy 1701 HisAlaGlnIleuAlaLeuGlyValIleSerGlyValIleProGlyValIleAlaIleLeuSerPheAsp 1720  
 Db 5101 CATGCAAGCTGGCACTGGGAGCCAGCGGGGGGCTGGGCAATGTGTGTGTGTGTGTGTGTGT 5160  
 Qy 1721 TyrGlyGlnGlnIleArgAlaGlnIleuLeuGlyProArgLeuIleGlyLeuIleIleuSer 1740  
 Db 5161 TATGGGAG 5220  
 Qy 1741 AsnIleThrValGlyValIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly 1760  
 Db 5221 AACATTAAGCAGCTGTTGGGTCAACAGAGGCTCCCAAGGGCCCATCAAGCAGAAAGTGGCT 5280  
 Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGlyGlnIleValAsnSerLeuAspProSer 1780  
 Db 5281 TGT 5340  
 Qy 1781 HisGlyGlnSerIleAsnValGlnGlnIleCysSerLeuProAspProCysAspSerAsn 1800  
 Db 5341 CATGGGAG 5400

Qy 1801 ProCysProAlaAsnSerIleCysSerAspAspThrAspSerIleSerCysSerCysAsp 1820  
 Db 5401 CCGTGTCT 5460  
 Qy 1821 ProGlyTyrTyrIleAspAsnCysThrAsnValCysAspLeuAsnProCysGlnIleGln 1840  
 Db 5461 CCAAGTACTATGT 5520  
 Qy 1841 SerValCysThrArgIleProSerAlaProIleGlyIleThrCysGlyCysProProAsn 1860  
 Db 5521 TGT 5580  
 Qy 1861 TyrLeuGlyProTyrCysGlnIleThrArgIleAspGlnProCysProAspGlyIleThrGly 1880  
 Db 5581 TACCTGGGCAATCTGT 5640  
 Qy 1881 HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn 1900  
 Db 5641 CATCCCAATGT 5700  
 Qy 1901 LysThrSerGlyGlyCysHisCysIleGlnIleThrArgProProGlyIleSerProThr 1920  
 Db 5701 AAGACAGCGGAGAGT 5760  
 Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyAsp 1940  
 Db 5761 TGCTCTGT 5820  
 Qy 1941 GlyGlnCysProCysIleProGlyValIleGlyValGlnIleCysAspArgCysAspAsnPro 1960  
 Db 5821 GGCAGAGT 5880  
 Qy 1961 PheAlaGlyValIleThrIleAsnGlyCysGlnValAsnIleThrAspSerCysProArgAlaIle 1980  
 Db 5881 TTGT 5940  
 Qy 1981 GlnAlaGlyIleThrIleProArgThrArgPheGlyLeuProAlaAlaIleProCysPro 2000  
 Db 5941 GAGGCTGT 6000  
 Qy 2001 LysGlySerPheGlyIleThrAlaValArgHisCysAspGlnIleValArgGlyIleProPro 2020  
 Db 6001 AAGAGCTCTTTGGGAGT 6060  
 Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuValGlyPheAlaGlnArgLeu 2040  
 Db 6061 AACCTTCACTGT 6120  
 Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnIleuAlaLeuLeuLeuArg 2060  
 Db 6121 CAGCGAATGT 6180  
 Qy 2061 AsnAlaThrGlnIleThrAlaGlyTyrPheGlySerAspValIleValIleValGlnLeu 2080  
 Db 6181 AACGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 6240  
 Qy 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 Db 6241 GCAACGGGCTGT 6260  
 Qy 2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
 Db 6301 GAGCTGCACTTCACTGAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6360  
 Qy 2121 LysArgHisThrGlyIleuValIleGlnIleThrGlnGlyValIleThrAlaIleThrLeuGlnHis 2140  
 Db 6361 AAGGGGCACTGT 6420  
 Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisIleThrTyrLeuSerProPhe 2160  
 Db 6421 TATGAGGCTATCGCAATGT 6480

2161 ThrValThrProLeuValIleSerValValArgLeuAspLysGlyAsnPheAla 2180  
6481 ACCATGTCACGGCCCAACATTCATCTCCGTAGTGGCTTGACAAGGAACTTGGCT 6540  
2181 GlyAlaLeuLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAsnLeuGluThr 2200  
6541 GGGGCCAAGTGGCCCGCTACAGAGCCCTGGTGGGAGAGCCGCCGACCTTGAGACA 6600  
2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
6601 AAGTCATCTGCTGAGTCTGCTTCACAGAGACCCGCCCTGGTCAAGGCCCGAGGC 6660  
2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGluArgArgHisProGluLeu 2240  
6661 CCGGAGAGGCCCAAGAGCAGAGAGCTGGCAGCGCAGCAGAGCGGACCCGAGAGCTG 6720  
2241 SerGlnGlyAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
6721 AGCCAGAGTGGCTGGCCAGCGTCATCATCTACCGCACCTGGCCGGCTTACCTGCT 6780  
2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
6781 CATACTATGACCTCTGACAGGCGACCTTGAGGTCCCAAGCCCGATCATCAACA 6840  
2281 ProValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLys 2300  
6841 CCGTGGTGAAGCATCGCTCCATGATGATGAGAGACTTCTGCCCGGCGCTGACAA 6900  
2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrLysProIleCysVal 2320  
6901 CCGTACAGGTGAGTCCGCTCGCTGAGACAGAGAGCGGACCAAGCCCATCTGTGTC 6960  
2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyGlyTyrTrpSerAlaArgGlyCysGlu 2340  
6961 TTCTGGAACATTCAATCCCTGGTCACTGGACAGGTGGCTGGTGGCCAGAGCTGTGA 7020  
2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
7021 GTCGTCTTCGCGATAGAGCCACGTACGCTGCCAATGCAACATGACGAGCTTCGCT 7080  
2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThr 2380  
7081 GAGCTCATGAGCTTCTCGCGGAGAAATGGGAGATCTGCTCACTGAAGACATCAACA 7140  
2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu 2400  
7141 TACGTGGCTCTAGGTCTCACTGGCTGCCCTTCTGCTCACTCTTCTTCTCACTCTC 7200  
2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
7201 TTGCGTATCTCGCTGCCCAACCAAGCGCATCCGATCACTGACACTGGCCCTGGGC 7260  
2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
7261 CTGGCTCAAGTGGTCTTCTCCCTGGGAAATCAACAGGCTGACTCCCTTTTGGCTGACA 7320  
2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460  
7321 GTCATTGGCATCTGCTGCACTCTCTGACCTTCTGACCTTTCCCGGGCTCTGCTGAG 7380  
2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
7381 GCGTTGACACTGATACCGGCACTCACTGAGTGGCGGATGTCAACACCGGCCCATAGCCC 7440  
2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
7441 TTCTACTACATGCTGGGCTGGGCGTGGCTGCTTATCATCAAGGCTTACCGTGGGCTTG 7500  
2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520  
7501 GACCCCGAGGGCTACGGGAACCTGACTTCTGCTGCTCCATCATATGACAGGCTCATC 7560  
2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540

7561 TGGAGTTTGTGGCCCGGCTGGCTTTGGCGTCCGATGAGTGTCTTCTGTAACATCTCG 7620  
2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLysGlyGlyProValSer 2560  
7621 GCGGCCCGGGCTCTGTGGTGGCCAGCGGCGGCTTTGAAAGAAAGATCTGTCTCG 7680  
2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
7681 GGCCTGAGCCCTCTTCCCGCTCTCTGCTGCTGAGGCCACGTGGTGTGCTGACCTG 7740  
2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
7741 CTCTGTCAACAGGACACCTCTCTTCCACTACCTTCTTGTGTAACCTGCAATTGCATC 7800  
2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGluValArgValAlaLeu 2620  
7801 CAGGGCCCTTCAATCTTCTCTCTATGTGGTCTTACAGAGAGTCCGGAACCACTC 7860  
2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
7861 AAGCTTGGCTGAGCGCGAGGCCAGCCCTGACCTTGTCTGACCAACAGTCCACCTG 7920  
2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
7921 ACCTGCTCTCAACTGCCCCAGCCCCCTACGCAATGGGCGGCTGTACAGGCCCTACGGA 7980  
2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGluProSerTyrIle 2680  
7981 GACTGGCGGCTCTCTGACAGCACAGTGTCTGGCGAAGATGACCGCCATCAATC 8040  
2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
8041 CCTTCTTGTGAGGAGAGAGTCCGCACTGAACCTTGCCCAAGGCCCTTGGCTGGGG 8100  
2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
8101 GATCCAGGAGCTGTCTCTGAGAGTCAAGACAGAGCATGATCTGACACAGCATCC 8160  
2721 AspSerAspLysSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
8161 GACAGTACCTGTCTTGAAGACACAGAGTGTCTCTTATGCTCTTACCACTCATCA 8220  
2741 AspSerGluGluGluGluGluGluGluGluGluGluAlaPheProGlyGluGluGly 2760  
8221 GACAGTGAAGGAAGGAAGAGAGAGAGAGAGAGGCGCTTCCCTGAGAGAGAGGCG 8280  
2761 TrpAspSerLeuLeuGlyProGlyAlaGluArgLeuProLeuHisSerThrProLysAsp 2780  
8281 TGGATACCTGCTGGGCGCTGGAGCAGAGAGACTGCCCTGCAAGTATCTCCAAAGAT 8340  
2781 GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800  
8341 GGGGGCCAGGGCTGGGAGGCCCCCTGGCCAGAGACTTTGGAGCAACAGCAAAAGAG 8400  
2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
8401 AGTATGGCAACGGGGCCCTGAGAGGCGCTGGGAGAAATGAGAGATGCCCTGTCTGA 8460  
2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyIleLeuLys 2840  
8461 GAGGGGTCCCTAAGGCCCCCTTCAAGGCTCTTCTGCTCCAGGCTCAACAAAGCATCTTAAG 8520  
2841 LysLysCysLeuProThrIleSerGluLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
8521 AAGAGTGTCTGCCACATACAGAGAGAGAGAGAGCTCTCTGCGCTCCCTTGAGACA 8580  
2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
8581 TGCAAGAGGTCTTCCCGGGGCTCTCCGCTAGTAGGCAAGCGGGGCGGCCCCCTCCC 8640  
2881 ArgProProProArgGlnSerLeuGlnGluGluLeuAsnGlyValMetProIleAlaMet 2900

Db 8641 CGCCACGCGCCGCGACAGCTTCAGAGACGCTGAACGCGGTCTCATGCCATGCCATG 8700  
 Qy 2901 SerIleuValaGlyThrValaAspGluAspSerSerGlySerGluPheLeuPhePheAsn 2320  
 Db 8701 AGCATCAAGGACGACGAGTGAATGAGACATCTGTCAGAGCTCCGAATTTCTTCTTAAC 8760  
 Qy 2921 PheLeuHis 2923  
 Db 8761 TTCCTGCAT 8769

## RESULT 2

US-10-225-567A-523  
 ; Sequence 523, Application US/10225567A  
 ; Publication No. US20030113798A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lifespan Biosciences  
 ; APPLICANT: Brown, Joseph P.  
 ; APPLICANT: Burner, Glenna C.  
 ; APPLICANT: Roush, Christine L.  
 ; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
 ; FILE REFERENCE: 1920-4-4  
 ; CURRENT APPLICATION NUMBER: US/10/225,567A  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: 60/257,144  
 ; PRIOR FILING DATE: 2000-12-19  
 ; NUMBER OF SEQ ID NOS: 2292  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 523  
 ; LENGTH: 10531  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-225-567A-523

## Alignment Scores:

Pred. No.: 0 Length: 10531  
 Score: 15545.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 15 Gaps: 0

US-09-916-849A-3 (1-2923) x US-10-225-567A-523 (1-10531)

Qy 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 63 ATGGAGAGCCGCGCACCGGCTCCCTCCCAAGCGCGCGCGCTGCTGCTG 122  
 Qy 21 LeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProGlyAspSerLeu 40  
 Db 123 TTGCTGCTGCTGCTGCGCGCGCTATTGGAGACCAAGTGGGCGCTGCTGCTG 182  
 Qy 41 G1SerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyThrLeuGlyProSer 60  
 Db 183 GGGTCCAGGGACGAGGCTCTTCGGGGGCTTGGCCCCCATGGGCTGCTGCTGCTG 242  
 Qy 61 SerAlaSerAsnLeuThrLeuGlyThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 243 TCACGCTCGAACCTCTGGCTCTTACCAAGCGCTGAGGATGGGGGACATGAGCTGACT 302  
 Qy 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGluHis 100  
 Db 303 GGGCAGCTGTAACCCCAACAGATGGCTGAGGGTTGGTGTCCAGATCCGAGGCCCAT 362  
 Qy 101 I1ProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
 Db 363 ATTCCTTACCAACAGCTCTTGAAGGCTGCGCTGGAGCTCTCCCTCTGAGGATGGA 422  
 Qy 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGluGluHisProCysLeuVala 140  
 Db 423 GGGCAGCTTTCCTCCACAGGGGACAGCTCACTGCCCCAGAGGACCCGCTTAAAGGCT 482  
 Qy 141 ProArgLeuArgCysGlnSerCysValLeuAlaGlnAlaProGlyLeuAlaGlyGlu 160

Db 483 CCAGGCTCAGATGCCAGTCTGCAAGCTGAGCAGAGCCCGGCTCAGGCGAGGAGA 542  
 Qy 161 ArgSerProGluGluSerLeuGlyValArgArgValArgAsnValAsnThrAlaProGln 180  
 Db 543 AGGTCAACCAAGAGTCCCTGGGCTGGCGCTCGAAGAAAGAAATGAATACAGCCCCCAG 602  
 Qy 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db 603 TTCAGCCCCCAAGCTACAGGCGCACAGTCCGAGAACCAAGCAGACAGGACCCCTGTT 662  
 Qy 201 AlaSerLeuArgAlaIleAspProAspGluGlyValGlyArgLeuGlyTyrThrMet 220  
 Db 663 GCATCCCTTACAGGCGCTTACCCGAGCGGGTGGCGGCTGACTGAGTACACACAG 722  
 Qy 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 723 GATGCCCTTTGATATGCGCTTCAACAGATCTTCTCTTGGACCCAGTCACTGATGCA 782  
 Qy 241 ValThrThrAlaGluGluLeuAspArgGluThrTyrSerThrHisValPheArgValThr 260  
 Db 783 GTAACCAAGCGAGAGAGCTGGATGCTGAGACCAAGAGCACCGCTTCAAGGTCAG 842  
 Qy 261 AlaGlnAspHisGlyMetProArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 843 GCGCAGAGACCAAGCGCATGCCGACAGAGTGCCTGGCTACACTCACTCTTGTTACT 902  
 Qy 281 AspThrAsnAspHisAspProValPheGluGlnGlnGlyTyrTyrLeuGluSerLeuArgGlu 300  
 Db 903 GACACCAATGACCTATACCTGTGTTCAGACAGAGATACAGAGAGCCCTCAGGGAG 962  
 Qy 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 963 AACCTGAGGTTGGCTATAGAGGCTCACTGTCAAGGCGACGAGATGATGATGCCCTCCC 1022  
 Qy 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyValSerProSerGlyValPhe 340  
 Db 1023 AATGCCAATATTCGTATACCGCTGTGAGGGGCTCGGGGCGCCCTCTGAAGTCTT 1082  
 Qy 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 1083 GAGATGACCTCTGCTGCTGGGTGATCCGACCCCGTGTGATGGGAGAGAGGTG 1142  
 Qy 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1143 GAATCTTACAGCTGACGCTGATAGGAGGACAGTGAACAGGGTGGGACCCCGGCTTCCGAG 1202  
 Qy 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400  
 Db 1203 ACCACAGCGCGTGTCTTCTCTGTGAGAGATGACAAATGATATATCCCCCACTTAAG 1262  
 Qy 401 GluValArgTyrValValGlnValArgGluAspValThrProGlyValaProValLeuArg 420  
 Db 1263 GAGAACCGCTATGTGTCACAGTGGAGAGATGTACTCCAGGGGCCCAAGTACTCCA 1322  
 Qy 421 ValThrAlaSerAspArgAspGlySerAsnAlaValValHisTyrSerIleMetSer 440  
 Db 1323 GTCAAGCTCTGGATGACAAAGAGGACCAATGCGGTGTGCACTATGCAATCAAGAT 1382  
 Qy 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAspValValSer 460  
 Db 1383 GGCATATGCTGGGAGACATTTATCTGAGATGCCCAAGCTGAGCTGTGATGTGTAGC 1442  
 Qy 461 ProLeuAspTyrGluThrThrValArgGluTyrThrLeuArgValaArgAlaGlnAspGly 480  
 Db 1443 CTTCTTGATATGACGACCAAGAGATCACCTTACGGGTTGAGACACAGATGTGGC 1502  
 Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1503 CGTCCCACTCTCTATGTCTCTGCTTGTGACGTACAGTCTCTGATATCAACAGAC 1562  
 Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 1563 AATGCCCAATCTTGTGTACAGACCCCTTTCAGGCTACTGTCTGAGAGAGCGTCCCTTA 1622





Db	3783	ATGCGCTGCGGTCCGCTGCTGCGCTTCACTCTCCGCGCCCTTCATGCGCTCTCTCC	3842
OY	1261	ValLeuPheArgProIleHisProValGlyLeuArgCysArgCysProProGlyPhe	1280
Db	3843	GTGCTCTTCCGGCCCATCCACCCCGTCGAGAGGCTGCGCTGCGCTGCCGCCGCTTC	3902
OY	1281	ThrGlyAspArgCysGluThrGlyValAlaLeuLeuSerArgCysProCysGlyProHis	1300
Db	3903	ACGGGTGACTATGCGAGACCGAGGTGACCTGTCTACTCGCGGCTCTGTGGCCCAAC	3962
OY	1301	GlyArgCysArgSerArgGlyGlyGlyThrCysLeuCysArgAspGlyTyrThrGly	1320
Db	3963	GGGCGCTGCGCGAGCCGAGGAGGCGGCTACCTGCGCTGTGCGTGAAGCTTACCGGCT	4022
OY	1321	GluHisCysGlyValSerAlaArgSerGlyArgCysThrProGlyValCysLeuAsnGly	1340
Db	4023	GAGCACTGTGAGTGAAGTCTGCTGCACAGCCGTTGCACCCGGGTGTGCAGAAATGGG	4082
OY	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe	1360
Db	4083	GGCACTGTGTCAACTGCTGCTGGGCGGTTCAAGTGCAGTGCCTCTGGAGACTTC	4142
OY	1361	GluIysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380
Db	4143	GAGAGAGCCCTACTGCGCAGGTGACACCGCGAGCTTCCCGGCCACTCTTCATCACTTT	4202
OY	1381	ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIysGlnArg	1400
Db	4203	CGCGGCTGCGCGACAGCTTTTCATCTCACTCCGCGCTCTGTTGGCCAGAAAGAGGC	4262
OY	1401	AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlySerHisAspPheValAlaLeuGlu	1420
Db	4263	GACGGGTTGCTGTTGTATCAATGGCGTTTCATATGAGAGCATGACTTTGTGGCCCTGAG	4322
OY	1421	ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLeuSerThrThrValSer	1440
Db	4323	GTGATCCAGAGACAGTCCAGTCACTTCTCTGAGGAGAGTCAACCAACAGGTGTCC	4382
OY	1441	ProPheValProGlyGlyValSerAspGlyGlnThrHisThrValGlnLeuIysTyr	1460
Db	4383	CCATTGCGGCCCGAGGAGTCAATGGCAGTGGCATTAAGTGTACGCTGAATATCATC	4442
OY	1461	AsnIysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnIysValAla	1480
Db	4443	AATAGGCCACTGTGGGTCAAGCAGGGCTCCACAGGGCCCATCAAGACAAAGGTGCT	4502
OY	1481	ValValThrValAspGlyCysAspThrGlyValAlaLeuAspPheGlySerValLeuGly	1500
Db	4503	GTGGTGACCGTGATGCGCTGTGACACAGAGATGGCTTGGCTTGCATCTGTCTGGGC	4562
OY	1501	AsnTyrSerCysAlaAlaGlnGlyThrGlnGlySerIlyIysSerLeuAspLeuThr	1520
Db	4563	AACTACTCTGTGTGCGCCACAGGACCCAGAGGTGCACAAAGAGTCTGTGACTGAAG	4622
OY	1521	GlyProLeuLeuLeuGlyGlyValProAspLeuProGlyIysArgPheProValArgMetArg	1540
Db	4623	GGGCCCCGTACTAGAGCGGGGTGCTTACCTGCGCGAGCTTCCAGTCCGAATGGG	4682
OY	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
Db	4683	CAGTTCGTGGGTGTGATCGGAACTGTGAGGTGGACACCGGCACTAGACATGGCTGAC	4742
OY	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIlyIysValAsnValCysAspSer	1580
Db	4743	TTCAATGGCCAAAGAGGACCGGTGCTGCTGCTGTGCACAGAAAGATGTGTGACAGC	4802
OY	1581	AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTyrAspAlaPheSerCysGlyCys	1600
Db	4803	AACACTTGCAAAATGGGGGACCTTGGGTGAACCAAGTGGAGCGGTTCAAGTGGAGTGC	4862
OY	1601	ProLeuGlyPheGlyGlyIlySerCysAlaGlnGlyMetAlaAsnProGlnHisPheLeu	1620

Db	4863	CCCTTGGGCTTTGGGGGCGAAGACTGGGCCCGCAAGAAATGGCCAAATCCACAGCACTTCTCTG	4922
Oy	1621	GIySeSerSeuValA1aTPrHiEgLYLeuSerLeuProIISeSerGIuProTIPrYLeu	1640
Db	4923	GGCAGCAGCCCTGGGCTGGCTGGCAATGGCTCTCGCTGGCCCAATCTCCCAACCTGTGTACTTC	4982
Oy	1641	SeSeruSerPheAqgThraArgGlnA1aAspG1yValLeuLeuGlnA1aIleThraArgLY	1660
Db	4983	AGCTCAATGTTCCGACAGCGCCAGGCGGACGAGATGTTCTGTGCGAGGCCATACCAAGGGG	5042
Oy	1661	ArgSerThrIleThraLeuGlnLeuArgGlnLYhiAValMetLeuSerValGluGlyThr	1680
Db	5043	CGCAGCACCAATCAACCTTACAGCTACGAGAGGAGGCCACAGTGATGCTGAGCGTGAAGGACCA	5102
Oy	1681	GIyLeuGlnA1aSerSerLeuAqgLeuGlnLYuProG1yThraA1aAspAAspG1yAspTPrHiS	1700
Db	5103	GGGCTTAAAGGCTCTCTCTCTCCGTGTAGCGAAGCCGGGCGCAATGACGGGATCTGGAC	5162
Oy	1701	HiSA1aGlnLeuA1aLeuGlnA1aSerG1yG1yProG1yHiSA1aIleLeuSerPheAAsp	1720
Db	5163	CATCCACAGCGTGGGACCTGGGAGCCAGGGGGGGCTGGCCATGCACTTGTGTCTTGAT	5222
Oy	1721	TyG1yGlnGlnAArgA1aGluG1yAAsnLeuG1yProAArgLeuHiEgLYLeuHiLeuSer	1740
Db	5223	TATGGGCGACGAGAGGACGAGGGGCAACTGGGGCCCCGGGCTGCATGTCTGCAACTGACG	5282
Oy	1741	AsnIleThraValG1yG1yIleProG1yProA1aG1yG1yValA1aAspG1yPheAArgG1y	1760
Db	5283	AACATTAACGTGGCGGGAATACTGGGCGACGGCGGTGTGGCCGTGGCTTTGGGGG	5342
Oy	1761	CySeLeuGlnG1yValAArgValSerAAspThrProGluG1yValAAsnSerLeuAAspProSer	1780
Db	5343	TGTTTTCAGAGGTTGGCGGGTGAAGGATACGCAAGAGGGGGTTTAAACACTGGATCCAGC	5402
Oy	1781	HiEgLYLeuSerIleAAsnValGluGlnG1yCySeSerLeuProAAspProCyAAspSerAAsn	1800
Db	5403	CATGGGAGAGCAATCAACGAGAGCAAGGCTGTAGCTGGCTGACCTGTGACTCAAC	5462
Oy	1801	ProCyAAspProA1aAsnSerTYTCySeSerAAspTPrAAspSerTYTCySeSerCyAAsp	1820
Db	5463	CCGGTCTCTGTACACTACTATTGGACAGCAAGACTGGGACAGCTATTCTCTGGACGCTGAT	5522
Oy	1821	ProG1yTYTCyTYTGTyAAspAAspCYThraAsnValCyAAspLeuAAspProCyEgLYhiGln	1840
Db	5523	CCAGGTACTAATGTGTACAACCTGACTAATGTGTGTACCTGAACCCGGTGTAGACACAG	5582
Oy	1841	SeValCYThraGlyAsProSerA1aProHiEgLYThraCYGluCYAspProAAsn	1860
Db	5583	TCTGTGTATACCCGCAAGCCCAAGTGGCCCCCAATGCTATACCTGGAGTGTCCCCCAAT	5642
Oy	1861	TyTLeuG1yProTYTCyEgLYThraG1yIleAAspGlnProCyAAspArgLYTPrTPrG1y	1880
Db	5643	TACCTTGGGCAATCTGTGAGACCAAGATTTGACACAGCTTGTCCCCGGTGGTGTGGGA	5702
Oy	1881	HiAAspProThraCYEgLYProCYAAsnCYAAspValSerLYEgLYPheAAspProAAspCYAAsn	1900
Db	5703	CATCCCAACATGTGGCCCAATGCAACTGTGATGTCAGCAAAAGCTTTGACCCAGACTGCAC	5762
Oy	1901	LYaThraSerG1yGluCYhiSValYEgLYAAsnHiSAThraArgProProG1ySerProThr	1920
Db	5763	AAGACAAGCGGCGAGTCCCACTGCAGAAAGAACCACTAACGGCCCCCAGGAGGACGCCAC	5822
Oy	1921	CyAseLeuCYAAspProCYAspProThrdG1ySerLeuSerArgValCYAAspProGluAAsp	1940
Db	5823	TGCTCTTGTGTACTGTACCCCAAGGCTCTTGTTCAGAGTCTGTACCTGTAGAGAT	5882
Oy	1941	GIyGlnCYAspProCYAAspProG1yValIleG1yAArgGlnCYAAspArgCYAAspAAspPro	1960
Db	5883	GGCCAGTGTCAATCAAGCCAGAGGTGATCGGGGGGTCAAGTGTGACCCGCTGCAACAACCT	5942
Oy	1961	PheA1aGlnValThraThraAsnG1yCYEgLYValAAsnTYThraSerCYAspProA1aIle	1980
Db	5943	TTTCTCTAAGTCAACCAATAGCTGTGAAGTAAATATATGACATGCTCCACAGAGCAAT	6002

QY 1981 GlnAlaGlyTlLeTrrPProAlaGlyThrArgPheGlyLeuProAlaAlaProCysPro 2000  
DB 6003 GAGGCTGGGATCTGGTGGCCCGCTACCGGCTTCGGGCTGCTGCTGCTCCCTGCCCC 6062  
QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnLysIleAspGlyTrrPLeuProPro 2020  
DB 6063 AAGAGCTCTTGTGGAGCTGCTGGGCGCAGTGTATAGACACAGGGGGTGGCTCCCGCCA 6122  
QY 2021 AsnLeuPheAsnCysTrpSerLleThrPheSerGlyLeuLysGlyPheAlaGlnArgLeu 2040  
DB 6123 AACCTCTTCAAGTCAAGTCCATCACTTCTCAGAACTGAAAGGCTTCGCTGAGCGGCTA 6182  
QY 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
DB 6183 CAGGGGATGATGATGAGGCTGAGACTGAGGCGCTCCAGACAGCTAGCGCTGCTCTGCGC 6242  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTrrPheGlySerAspValLysValAlaTrrGlnLeu 2080  
DB 6243 AACGCGACGAGACACAGGCTGGCTACTTGGCGACGACGTCAGAGTGGCTTACAGCTG 6302  
QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyLeuSerAlaThrGln 2100  
DB 6303 GCCACGGGGTGGTGGCGCCACGAGACGCCAGGGGGCTTGGGCTGTCTGCCACACAG 6362  
QY 2101 AspValHisPheThrGlnAsnLeuLysValGlySerAlaLeuLeuAspThrAlaAsn 2120  
DB 6363 GAGGTGCACTTCACTGAGAACTGCTGCGGGTGGGACGCGCTCTGACACAGCGCAC 6422  
QY 2121 LysArgHisThrGlyLeuLysGlnGlnThrGlnGlyTrrAlaTrrPLeuLeuGlnHis 2140  
DB 6423 AAGGGGACCTGGAGGCTGATTCAGACACAGAGGGTGGACCGCTGGCTGCTCCAGCAC 6482  
QY 2141 TyrGlyLysAlaTrrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTrrLeuSerProPhe 2160  
DB 6483 TAGAGGCGCTACGCGAGTGGCTTGGCGCCAGACATGGGGACACCTACCAAGCCCTTC 6542  
QY 2161 ThrLleValThrProAsnLleValLleSerValValArgLeuAspLysGlyAsnPheAla 2180  
DB 6543 ACCATGCTCAAGCGCCACATTTGTCATCTCCGTAGTGGCTTGGACAAAGGAACTTTGCT 6602  
QY 2181 GlnAlaLysLeuProArgTrrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
DB 6603 GGGGCGCAGGTGGCCCGCTGACAGGCGCTGCTGGGAGAGCGCCCGGACCTTGGACA 6662  
QY 2201 ThrValLleLeuProGlySerValPheArgGlyThrProProValValArgProAlaGly 2220  
DB 6663 AAGTCACTTGGCTGAGTGTGTCTTCAAGAGACGCCCCGCTGTGACAGGCCCGCAGGC 6722  
QY 2221 ProGlyGlnAlaGlnGlnProGlyGlnGlnLeuAlaArgArgGlnArgArgHisProGlyLeu 2240  
DB 6723 CCGGAGAGGGCCAGGAGCAGAGAGGCTGGCACGGGACAGCGACGCGACCGGAGCTG 6782  
QY 2241 SerGlnGlyGlnAlaValAlaAspValLleLleTrrArgThrLeuAlaGlyLeuLeuPro 2260  
DB 6783 AGCGAGGGTATAGGCTGTGGCAGCGTCATCATTCAGCGACCTGTGGCGGGCTACAGCTC 6842  
QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProLleIleAsnThr 2280  
DB 6843 CAAACTATATGACCTTGAACAGCCAGCTTGAAGTCCCAAGGCCGATCAACAACACA 6902  
QY 2281 ProValLysSerLleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspLys 2300  
DB 6903 CCGGTGTGATGATCAGCGCTCAGATGATGAGAGAGTTCCTGCGCGGGCGCTGACAAA 6962  
QY 2301 ProValThrValGlnPheArgLeuLeuGlnThrGlnGlnArgHisPheProLleCysVal 2320  
DB 6963 CCGGTACGGGTGAGTTCGCGCTGAGAGACAGAGAGCGGACCAAGCCCATCTGTGTC 7022  
QY 2321 PheTrrAsnHisSerLleLeuValSerGlyThrGlyTrrGlyTrrSerAlaArgGlyCysGln 2340  
DB 7023 TTCTGGAAACATTCAATTCCTGTGATGATGAGCAGGTGGCTGTGCTGGCCAGAGGCTGTGA 7082

QY 2341 ValValPheArgAsnGlySerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
DB 7083 GTGCTCTTCGCAATGAGAGCGACGTCAGCTGCAATGAGAACCAATGACAGACTTGCT 7142  
QY 2361 ValLeuMetAspValSerArgArgGlnGlnGlnLleLeuProLeuLysThrLleThr 2380  
DB 7143 GTGCTATGATGACCTTCTCTGGCGGAGAAATGGGAGATCTGCACTGATGAGACATGACA 7202  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu 2400  
DB 7203 TAGGTGCTTAGGTATCACTTGGCTGGCTTCTCTCACTTCTTCTTCTCACTCTC 7262  
QY 2401 LeuArgLleLeuArgSerAsnGlnHisGlyLleArgArgAsnLeuThrAlaAlaLeuGly 2420  
DB 7263 TTGCGTATCTGGGCTCCAAACCAACGCGCATCCGATCACTGACAGCTGCGCTGGGC 7322  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyLleAsnGlnAlaAspLeuProPheAlaCysTrp 2440  
DB 7323 CTGGCTCAGCTGGTCTTCTCTCGGGATATCAACAGGCTGACCTCTTTCCTGTCACA 7382  
QY 2441 ValLleAlaLleLeuLeuHisPheLeuTrrLeuCysThrPheSerTrrAlaLeuLeuGln 2460  
DB 7383 GTCATTTGCCATCTGCTGACCTTCTGTACTCTGTGACCTTTTCCTGGGCTGCTGGAG 7442  
QY 2461 AlaLeuHisLeuTrrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480  
DB 7443 GCCTTGCACTGTATCCGGGCACTCACTGAGGTGGCGGATGTCAACACCGGCCCATGGCG 7502  
QY 2481 PheTrrTrrMetLeuGlyTrrGlyValProAlaPheLleThrGlyLeuAlaValGlyLeu 2500  
DB 7503 TTCTATTAATGCTGGGCTGGGCGGTGGCTGCTGCTTCATACAGGGCTACCGGGGCTG 7562  
QY 2501 AspProGlnGlyTrrGlyAsnProAspPheCysTrrPLeuSerLleTrrAspThrLeuLle 2520  
DB 7563 GACCCGAGGGCTACGGGAACTCTGACTTCTGCTGCTTCTCATCTTACAGAGCTCATTC 7622  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrrLleLeu 2540  
DB 7623 TGGAGTTTGTGGCGCGGCGGTGGCTTGGCGCTGATGAGTGTCTTCTGTATCATCTCG 7682  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLysLysGlyProValSer 2560  
DB 7683 GCGGCGCGGCGCTCTGAGTGTGCTCCAGCGGCGGCTTGGAAAGAAAGTCTGTCTCG 7742  
QY 2561 GlnLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrrPLeuLeuAlaLeu 2580  
DB 7743 GGCCTGACAGCTTCTTCCGCGCTCTCTGCTGCTGAGGCGCACGTTGGCTGCTGACCTG 7802  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTrrLeuPheAlaThrCysAsnCysLle 2600  
DB 7803 CTCTGTCAACAGGACACCTCTCTTCACTACCTCTTGTGATCACTGCAATTGCATTC 7862  
QY 2601 GlnGlyProPheLlePheLeuSerTrrValValLleSerLysGlnValArgLysAlaLeu 2620  
DB 7863 CAGGGCCCTTATATCTTCTCTCTATGTGTGCTTAAAGAGAGGTGCGGAAAGACATC 7922  
QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
DB 7923 AAGCTTGGCTGAGCGCGAGCGGACCGCTTACCTGCTCTTACACCAAGTCCACCTG 7982  
QY 2641 ThrSerSerTrrAsnCysProSerProTrrAlaAspGlyArgLeuTrrGlnProTrrGly 2660  
DB 7983 ACCCTGCTCTTACACTGCGCCCGCTTACGAGATGGGGCGGCTGTACACAGCTTACGGA 8042  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTrrLle 2680  
DB 8043 GACTGGCGGGCTCTCTGACAGACAGTGCCTGGGGAGAAAGTCAAGCTACATCATC 8102  
QY 2681 ProPheLeuLeuArgGlnGlySerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
DB 8103 CCTTCTTGTGAGGAGAGATCGGCACTGAAACCTTGGCGAAAGGCCCTTGGCTGGGG 8162  
QY 2701 AspProGlySerLeuPheLeuGlnGlyLysArgGlnGlnHisAspProAspThrAspSer 2720

```

Db      8163 GATCCAGGACGCTGTTCTCGAAGGTCAAGACGACGATGATCTCGAAGCAGCTCC      8222
Qy      2721 AapSerAapLeuSerLeuGluAapAapGluSerGlySerYrAlaSerThrHisSerSer 2740
Db      8223 GACAGTGAACCTGCTCTTGAAGACGACCAAGATGCTCTTATGCTCTTACCACTGATCA 8282
Qy      2741 AapSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760
Db      8283 GACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8342
Qy      2761 TrpAapSerLeuLeuGlyProGlyValAgluAargLeuProLeuHisSerThrProlyAap 2780
Db      8343 TGGGATACCTGCTGAGGCTGAGACGACGAGACTGCTGACAGACTGCTGACAGAGAT 8402
Qy      2781 G1yG1yProG1yProG1yYrAlaProTrpProG1yAapPheG1yThrAlaYrAlaG1u 2800
Db      8403 GGGGGCCCAAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8462
Qy      2801 SerSerG1yAamG1yAlaProGluGluAargLeuAargG1yAamG1yAapAlaLeuSerA 2820
Db      8463 AGTAGTGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8522
Qy      2821 GluGluSerLeuGlyProLeuProG1ySerSerAlaGlnProHisGlyG1yLeuGly 2840
Db      8523 GAGGGGCTCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8582
Qy      2841 LysLysCysLeuProThrTrlSerGluYrSerSerLeuLeuAargLeuProLeuGluGlu 2860
Db      8583 AAGAAGTCTCTGCCACATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8642
Qy      2861 CysThrGlySerSerAargGlySerSerAlaSerGluGlySerAargG1yG1yProPro 2880
Db      8643 TGCAAGAGGCTCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8702
Qy      2881 ArgProProAargGluSerLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 2900
Db      8703 GCGCCACGCGCCGCGAGAGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8762
Qy      2901 SerLeuYrAlaGlyThrValAapGluAapSerSerGlySerGluPheLeuPheAasn 2920
Db      8763 AGCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8822
Qy      2921 PheLeuHis 2923
Db      8823 TTCCTGCAT 8831

```

## RESULT 3

```

US-09-788-711A-1
; Sequence 1, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 8871
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-1

```

## Alignment Scores:

```

Pred. No.: 0 Length: 8871
Score: 15518.50 Matches: 2923
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 0
Query Match: 99.83% Indels: 33

```

```

DB: 9 Gaps: 1
US-09-916-849A-3 (1-2923) x US-09-788-711A-1 (1-8871)
Qy      1 MetLysSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
Db      1 ATGCGAGAGCCCGGACACGGGCTCCCTCCCAACGCGCGCGCGCTGCTGCTCTG 60
Qy      21 LeuLeuLeuLeuLeuProProProLeuLeuG1yAapG1yAlaG1yProCysAargSerLeu 40
Db      61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qy      41 GlySerAargG1yAargGlySerSerG1yAlaCysAlaProMetG1yTrpLeuCysProSer 60
Db      121 GGGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
Qy      61 SerLysSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
Db      181 TCAGCTGCGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Qy      81 G1yHisLeuValProHisHisAapG1yLeuAargValTrpCysProGluSerGluAlaHis 100
Db      241 GGGCACCCTGATACCCCAACGATGAGGCTGAGGCTTGGTGTCTCAGAAATCCAGAGGCCAT 300
Qy      101 IleProLeuProProAlaProGluGlyCysProTrpSerCysAargLeuLeuGlyLeGly 120
Db      301 ATTCCTCTTACCAACAGCTCTTGAAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Qy      121 G1yHisLeuSerProGluGluGlyYrYrLeuThrLeuProGluGluHisProCysLeuValAla 140
Db      361 GGGCACCCTTCCCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Qy      141 ProAargLeuAargCysGlnSerCysLeuAlaGlnAlaProGlyLeuAargAlaGlyGlu 160
Db      421 CCAAGGCTCAGATGCCAGTCTGCAACTGCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
Qy      161 AargSerProGluGluSerLeuGlyYrAargGlyYrAargValAaThrAlaProGln 180
Db      481 AGGTCAACAGAAAGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Qy      181 PheGlnProProSerTrpGlnAlaThrValProGluAamGlnProAlaGlyThrProVal 200
Db      541 TTCAGGCCCTCCAGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
Qy      201 AlaSerLeuAargAlaLeaProAapGluGlyGluAlaGlyAargLeuGlyTrpMet 220
Db      601 GCATCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Qy      221 AapAlaLeuPheAapSerAargSerAargGlnPhePheSerLeuAapProValThrGlyAla 240
Db      661 GATGCCCTCTTGTATGATCCGCTCCACAGATTCCTCCCTGAGCCACGATCACTGCTGCA 720
Qy      241 ValThrThrAlaGluGluLeuAapAargGluThrTrpYrSerThrHisValPheAargValThr 260
Db      721 GTATACCAAGCCAGAGAGGAGTGAATGTAAGACAAAGAGCAACCAAGTCTTCAAGGATACG 780
Qy      261 AlaGlnAapHisGlyMetProAargAargSerAlaLeuAlaThrLeuThrTrlLeuValThr 280
Db      781 GCGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Qy      281 AapThrAasnAapHisAapProValPheGluGluGluGluGluGluGluGluGluGluGlu 300
Db      841 GACACCAATGACCAATGACCTGTGTGTTGACACGAGGAGGATCAAGAGAGAGGAGGAGGAGGAGGAG 900
Qy      301 AasnLeuGluValGlyYrGluValLeuThrValAargAlaThrAapG1yAapAlaProPro 320
Db      901 AACCTGAGAGGTGGCTATGAGGCTCATCTGACGGCCACGAGATGTATGATGCCCTCTCC 960
Qy      321 AasnAlaasnIleLeuYrAargLeuLeuGluGlySerGlyYrGlySerProSerGluValPhe 340
Db      961 AATGCCAATATTTCTGTATCCCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Qy      341 GluIleAapProAargSerGlyValIleAargThrAargGlyProValAapAargGluGluVal 360

```

```

Db 1021 GAGATCGACCTCGCTGGGGTGAATCCGAACCCCTGGCCCTGGATCGGGAGAGAGTGT 1080
Qy GluSerTyrGlnLeuThrValGluAlaSerAspGlnIYarAspProGlyProArgSer 380
Db 1081 GAATCTTCACGAGTACCGGTAGAGCAATGACAGAGGTCCGAGCCCGGGTCTCGAGAT 1140
Qy ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400
Db 1141 ACCACAGCCGCTGTTTCTTCTGTGTGAGATGACATGATTAATGCCCCCAAGTTAAT 1200
Qy GluIlyAspArgTyrValGlnValArgGluAspValThrProGlnAlaProValLeuArg 420
Db 1201 GAGAGGCGCTATGAGTCCAGGTGAGGAGATGATGATCCAGGGGCGCCAGTACTCCGA 1260
Qy ValThrAlaSerAspArgAspIlySerAspAlaValAlaIlyTyrSerIleMetSer 440
Db 1261 GTCAACGCTCGGATTCGAGACAGAGGGAGCAATGCCGTGAGTCACTATGACATGAGAT 1320
Qy GluAlaAlaArgGluGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
Db 1321 GGCAATGCTCGGGACAGTTTATCTGATGCGCACTGCAAGCTGTGATGTGTGAGC 1380
Qy ProLeuAspTyrGlnThrThrIlyGluIlyTyrThrLeuArgValArgAlaGlnAspGly 480
Db 1381 CCTTTGACTATGAGACAGACAGAGATGACCCCTACGGGTGCGACACAGATGTGTGC 1440
Qy ArgProProLeuSerAspValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500
Db 1441 CGTCCCACTCTATATGCTTGTGCTGTGATGACAGTACAGTCTCGATATACAGAC 1500
Qy AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520
Db 1501 AATCCCCCACTTCTGTAGACACCCCTTTCAGGCTACTGCTCGAGAGAGTCCCTTA 1560
Qy GlyTyrIleValLeuAlaValGlnAlaAlaIleAspAlaAspAlaGlyAspAspAlaArgLeu 540
Db 1561 GGGTACCTGTGCTTCAGTGTCCAGGTATGACAGCTATGCTGTGTGCAATGCCCGCTG 1620
Qy GluTyrArgLeuAlaGlyValGlyIlyIleAspPheProPheThrIleAsnAspGlyThrGly 560
Db 1621 GAATACCGCTTGTGCTGGGTGGACATGACTTCCCTTACACATCAACATGAGCAAGGC 1680
Qy TyrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db 1681 TGGATCTCTGTGGCTGGAATGGAAGTGAAGGAGAGTGAATTTCTACAGCTTGGGTA 1740
Qy GluAlaArgAspAlaGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1741 GAAGCTCGAGACCAATGCACTCCAGCACTCACTGCTCGGCAAGTGTCAAGCTGATGTC 1800
Qy LeuAspValAsnAspAspAspProThrPheThrGlnProGluIlyTyrThrValArgLeuAsn 620
Db 1801 CTGATGTCAACGACACATCACTTACCTTACCAACAGATGACAGTGGCGCTCAAT 1860
Qy GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db 1861 GAGGATCAAGCTGTGGGACCAAGGTGTGAAGGTGTGAGTGTGACCGGTGAATGCTCAT 1920
Qy SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer 660
Db 1921 AGTGTCACTACCTTACAGATCAAGTGGCAATCTGAAACCCCTTCTCATCAACAGC 1980
Qy GlnSerGlyGlyIlyLeuValSerIleuAlaLeuProLeuAspTyrIlyLeuGluAlaArgGln 680
Db 1981 CAAGTGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
Qy TyrValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db 2041 TATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2100
Qy AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 2101 AATGTACCGACCGCAACACCACTGCTGTGCTTTCAGAGCTCCCACTATACATGAT 2160
Qy ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db 2161 GTTATATAGAGACCGCGGGAGGACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
Qy ThrGlyIleuAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACAGTATAGAAATCCCGCATCACTACTTCAATGAGAGACAGATCCCCAGTTCCGATC 2280
Qy AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db 2281 GATGACAGACCGGGGGGTGTCAACACCGCTGAGCTGTGATACAGAAAGCAAGTGTCT 2340
Qy TyrThrLeuAlaIleThrAlaArgAspAspGlyIleProGlnIlySerAspThrThrTyr 800
Db 2341 TACACCTTGGCATTTACTGTCTCGAGACAAAGGCAATCCCAAGAAAGTCCGACACCTAC 2400
Qy LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db 2401 CTGGAGATCTGTGTGAACAGCGTGAATGACAAATGCCCTCAAGTTCTGGGAGCTCTAC 2460
Qy GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGGCAGTGTCTATAGAGATGTGCCACCTTCACTAGCTCTGTGAGATCTCAGCCACT 2520
Qy AspArgAspSerGlyLeuAsnGlyIlyArgValPheTyrThrPheGlnGlyIlyAspAspGly 860
Db 2521 GATCGATTTCTGACTTATATGCAAGGTCTTTTCACTTCCAAAGAGCGACAGATGA 2580
Qy AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GACGGTGACTTATATGTGTGAGTCACTGATGATGATGATGATGATGATGATGATGAT 2640
Qy ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspIlyGlyMetProPro 900
Db 2641 CAGAGAAAGGTGCCAGATATGCTTGTGGGCAATGACAGTGAACAGGGGAGTCCCA 2700
Qy AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspProVal 920
Db 2701 GCCCGACACTTATGAGATGACAGTCACTGTGTGTGATGTGATGATGATGATGATGAT 2760
Qy PheGlnGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTTGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2820
Qy AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db 2821 GCCCGGGTCAAGCCCACTGACCCCGATGAAAGGCAACAAATGCCAGATGATGATGAT 2880
Qy ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 2881 GTGAGGGCAATCTCTGAGGTCTTCCAGCTGAGACATCTTCTCCGGGAGCTGACAGCC 2940
Qy LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
Qy AlaProLeuValSerArgAlaThrValAlaValArgLeuLeuAspArgAsnAspAspPro 1020
Db 3001 GCTTCTGTGTGAGCGGGGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 3060
Qy ProValLeuGlyAsnPheGluIleLeuPheAsnAspTyrValThrAsnArgSerSer 1040
Db 3061 CAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3100
Qy PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3121 TTTCCCTGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
Qy ThrTyrSerPheGluArgGlyIlyAsnGlyIlySerLeuValLeuLeuAsnAlaSerThrGly 1080
Db 3181 ACTTACAGCTTGTGAGGGGAAATGAACTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 3240

```



OY	1061	GIuIEuLyLEuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	3241	GAGCTGAAGCTTAAGCCGGCCACTGGACAAACACGGGCTCTGGAGGCCATCATGAGCTGG	3300
OY	1101	LeuValSerAspArgLyValHisSerValThrAlaGlnCyValaLeuArgValThrIleIle	1120
Db	3301	CTGGGTCAAGACGGGTGACACAGCTGACCCGCCAGTCCGGCTGGCTGATCAATCATC	3360
OY	1121	ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg	1140
Db	3361	ACCGATGAGATGCTCAACCAACAGCATCAAGCTGGCCCTGGAGAGCATTCACCGAGGCG	3420
OY	1141	PheIleuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaThrLeuAlaThrPro	1160
Db	3421	TTCCGTGACCACTGCTCAAGGCTCTTCACTTCACTCAAGGGGTGGCGCCGACCTGGACGGCA	3480
OY	1161	ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle	1180
Db	3481	CCGAGCAACGTGTGTGCTTCAAGCTACAGGGAGACACGACGCCCCGGGGGGCCACATC	3540
OY	1181	LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyValProProPheLeu	1200
Db	3541	CTCAACGTGAGCTGTCCGGTGGGCGACGCGACAGGGCCCGGGGGGGCGGCCCTTCTCG	3600
OY	1201	ProSerGluAspLeuGlnGluArgLeuTyriLeuAsnArgSerLeuLeuThrAlaIleSer	1220
Db	3601	CCCTCTGAGGACCTGACAGAGGGCTTAATCTCAACCCGACCTGCTGACCGGCATCTCG	3660
OY	1221	AlaGlnArgValLeuProPheAspAspAsnIleCyValaArgGluProCyGluAsnThr	1240
Db	3661	GCAACGCGCGCTGCTCCCTTGGACCAACAATCTGCTCGCGAGACCTTGGAGAACTAC	3720
OY	1241	MetArgCyValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260
Db	3721	ATGCGCTCGAGTCCGATGCTGGCTGTCACTCTCCGCGCCCTTATGAGCTCTCCCTCC	3780
OY	1261	ValLeuPheArgProIleHisProValIleGlyLeuAlaGlyValArgCyAspProProGlyPhe	1280
Db	3781	GTGCTCTTCCGGCCCATCAACCCCTCGAAGGGCTGGCTCCGCTGCCCCCGGCTTC	3840
OY	1281	ThrGlyAspTyiCyGluThrGluValAspLeuCyTyiSerArgProCyGlyProHis	1300
Db	3841	ACGGGTGACTACTGGAGAACCGAGGTGAACCTCTGCTACTGCGGCCCTGTGGCCCCAC	3900
OY	1301	GlyArgCyAspSerArgGluGlyGlyTyiThrCyIleuCyAspArgGlyTyiThrGly	1320
Db	3901	GGGCGCTCCCGACGCGGAGGGCGGCTCAACCTGCTCTGCTGAGAGGCTAACAGGGT	3960
OY	1321	GluHisCyGluValSerAlaArgSerGlyArgCyThrProGlyValCyValAspAsnGly	1340
Db	3961	GAGCACTGTAGGTAGTGTCTGTCAAGACGCTTGACCCCGGGGTGTGCAAGAAATGGG	4020
OY	1341	GlyThrCyValAsnLeuLeuValGlyGlyPheIleCyAspCyProSerGlyAspPhe	1360
Db	4021	GGCACCTGTGTCAACTGCTGTGGGGCGGTTCAAGTGCATGTGCCATCTGGAGACTTC	4080
OY	1361	GlyIleProTyiCyGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380
Db	4081	GAGAGCCCTTACTGTCAGGTGACACGCGCAGCTTCCCGGCCACTCTTCATCACTTT	4140
OY	1381	ArgGlyLeuAspArgAlaArgPheHisPheThrLeuAlaLeuSerPheAlaThrTyIleGluArg	1400
Db	4141	CGCGGCTCGCGCACGCTTTCACCTTCACTTCACTCGGCTCTGTGTTTGCACAAAGAGCGC	4200
OY	1401	AspGlyLeuLeuLeuTyiAsnGlyArgPheAsnGlyuHisAspPheValAlaLeuGlu	1420
Db	4201	GACGGGTGTCTGTGTACAAATGGCGGTTCAATGAGAGCAATGACTTGTGTGCTCGAG	4260
OY	1421	ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGlyLeuSerThrThrValSer	1440
Db	4261	GTGATTCAGAGACAGGTCCAGCTCATCTTCTGTGAGGGAGTCAACCAACAGGTGTCC	4320

QY	1441	ProPheValProGlyValAspArgGlyGlnTrpHisThrValGlnIleuValPheTyr	1460
Db	4321	CCATTTCGGCCCGGAGAGCTCAGTAAGAGGCGCAGTGGCACTACGGTGGAAATCTTAC	4380
QY	1461	AsnAspProIleuLeuGlyGlnThrGlyLeuPProGlnGlyProSerGlnGlnValAla	1480
Db	4381	AATAAGCACTGTTGGGTCAAGACAGGGCTCCACAGGGCCCATCAAGACAGAAAGTGGCT	4440
QY	1481	ValValThrValAspGlyCyAspThrGlyValAlaLeuArgPheGlySerValLeuGly	1500
Db	4441	GTGGTGAACCGTGAATGGCTGTGACACAGAGATGGCTTGGCTTCGGATCTGCTCGGGC	4500
QY	1501	AsnTyrSerCysAlaAlaGlnGlyLysThrGlnGlyGlySerIleuSlySerIleuAspLeuThr	1520
Db	4501	AACACTCTCTGTGCTGCCCCAGGGGACCCAGGGGTGGACAGAAAGTCTCTGAGACTGACG	4560
QY	1521	GlyPProIleuLeuLeuGlyGlyValProAspLeuProIleuSerPheProValArgMetArg	1540
Db	4561	GGGGCCCCGTACTAGCGGGGGTCTCTGACCTGGCCGAGAGACTTCCAGTCCGAATCGG	4620
QY	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
Db	4621	CAGTTCGGGGCTGCATAGCGGAACTCAGGTGACAGCGGCAATAGACATGCTGAC	4680
QY	1561	PheIleAlaAsnAsnGlyLysThrValProGlyCysProAlaIleSlyAsnValCyAspSer	1580
Db	4681	TTCAATTGGCAACATGGGACCGGTGCTGGTGGCTGCCCTCCAAAGAAAGATGTGTGACAGC	4740
QY	1581	AsnThrCysHisAlaAsnGlyGlyLysCysValAlaAsnGlnTrpAspAlaPheSerCysGlnCys	1600
Db	4741	AACACTGGCCCAATAGGGGGGCACTGGGTGAACAGAGGGACCGGTTCAGCTGGAGATGC	4800
QY	1601	ProIleuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu	1620
Db	4801	CCCCGTGGCTTTGGGGGGCAAGACTGGGCCAGAAATGGCCATTCACAGCACTTCTCG	4860
QY	1621	GlySerSerIleuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTyrTyrIleu	1640
Db	4861	GGCAGACAGCCCTGGGCTGGCAAGGCTCTCGCTGCCCATCTCCAAACCTGGTACTCTC	4920
QY	1641	SerIleuSerPheArgTrpArgGlnAlaAspGlyValIleuLeuGlnAlaIleThrArgGly	1660
Db	4921	AGCTCATGTTCCGACGCGCCAGGCGGACGAGATCTCTGCTGAGGCCATACAGGGGG	4980
QY	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGlyLys	1680
Db	4981	CGCAGCACCAATCACCTTACAGCTTACGAGAGGGCCAGCATGATGCTGAGGTGAAGGCACA	5040
QY	1681	GlyLeuGlnAlaSerSerIleuArgLeuGlnLysProGlyArgAlaAsnAspGlyAspTrpHis	1700
Db	5041	GGGCTTCAGGGCTCTCTCTCCGTCTGAGGACAGGGCGGGCCAAATGACGGTGACTGGCAC	5100
QY	1701	HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp	1720
Db	5101	CATGCACAGCTGGGACCTGGAGGCCAGGGGGGGCTGGCATGCACTTCTGTCTTGAT	5160
QY	1721	TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgIleuHisGlyLeuHisLeuSer	1740
Db	5161	TATGGGCGCAGAGAGCAGAGGGGCAACTGGGGCCCCGGCTGCATGTGCTGCACCTTAGC	5220
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly	1760
Db	5221	AACATAACATGGGGCGAATTCTGGGGCCAGCCGGCGGTGTGGCCGTGGCTTTCGGGGC	5280
QY	1761	CysIleuGlnGlyValArgValSerAspThrProGlnGlyValaAsnSerLeuAspProSer	1780
Db	5281	TGTTTGCAGGGGTGTGGCGGTGAAGGAATACGCACAGAGGGGGTTTAAACACTGGAATCCAGC	5340
QY	1781	HisGlyGlnSerIleAsnValGlnGlnGlyCysSerIleuProAspProCysAspSerAsn	1800
Db	5341	CATGGGGAGCAATCAACGTGAACAAAGGCTGTAAGCTTGACCTTGTGACTCAAAAC	5400
QY	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp	1820



Db	5401	CCGGTCTCTGTAACAGCTATTTGACAGACGAGCTGAGCAAGCTTCTGAGCTGGAT	5460	Db	6481	ACCATGTCACGCCCAACATTTGTCACTCCGTAAGTCGCTTGACAAAGGAATTGGCT	6540
Qy	1821	ProG1YTYTYrG1YAAspAenCyEThrAsnValCyAspLeuAAspProCySG1uH1AG1n	1840	Qy	2181	GLYAlaLeuLeuProArgTyrG1uAlaLeuArgG1G1uG1nProProAspLeuG1uThr	2200
Db	5461	CGAGGTACTAGTGTGACACTGACACTGACACTGACACTGACACTGACACTGACACTG	5520	Db	6541	GGGGCCAGCTGCCCCGCTACAGAGCCCTGCGGGGAGAGAGCCCCCGGACCTTGAGACA	6600
Qy	1841	SerValCyEThrArgLysProSerAlaProH1AG1YTYTYrG1uG1uG1uG1uG1uG1u	1860	Qy	2201	ThrVal1LeuLeuProG1uSerValPheArgG1uThrProProValValArgProAlaG1y	2220
Db	5521	TCTGTGTGTATCCCGCAAGCCAGAGTGGCCCCCATGAGTATACCTGAGAGTCCCCCAAT	5580	Db	6601	ACAGTCACTTGTGCTGAGTCTGTCTTCAAGAGAGCCCCCGTGTGAGGCCCGAGGC	6660
Qy	1861	TyrLeuG1yProTyrCySG1uThrArgG1uAspG1nProCySProAlaG1yTYTPTTGG1y	1880	Qy	2221	ProG1yG1uAlaG1nG1uProG1yG1uLeuAlaArgArgG1nArgG1nAspProG1uLeu	2240
Db	5581	TACCTTGGCCATACCTGAGACCGAGATTCAGACCTTGTCCCTGTGCTGTGGGGA	5640	Db	6661	CCCGAGAGGCCAGAGCCAGAGGAGCTGCGACCGCGACAGCGAGCGGACCCGGAGCTG	6720
Qy	1881	HisProThrCySG1yProCySAsnCyAspValSerLysG1yPheAspProAspCySAsn	1900	Qy	2241	SerG1nG1yG1uAlaValAlaSerVal1Le1Le1YTYTYrG1uThrLeuAlaG1yLeuLeuPro	2260
Db	5641	CATCCCAATGTGGCCCATCATCATCTGATGTCTGACAAAGGCTTTGACCCAGACTGCAC	5700	Db	6721	AGCCAGGGTGAAGCTTGGCCAGGATCACTTACCGCACCTTGGCCGGGCTACTGCTC	6780
Qy	1901	LysThrSerG1yG1uCySAsnCySAspLysG1uAsnH1sTYTYrArgProProG1ySerProThr	1920	Qy	2261	HisAsnTYrAspProAspLysArgSerLeuArgValProLysArgPro1Le1LeaAnThr	2280
Db	5701	AAAGCAAGCCGCGAGTCCACTGCAAGAGAACCACTACCGGCCCCCAGGCAAGCCCCACC	5760	Db	6781	CATPACTATGACCTTGACAAAGCGACGCTTGAGAGTCCCAAGGCCGATCATCAACA	6840
Qy	1921	CysLeuLeuCySAspCySAspProProThrG1ySerLeuSerArgValCySAspProG1uAsp	1940	Qy	2281	ProVal1Ser1LeSerValH1sAspArgG1uG1uLeuLeuProArgAlaLeuAspLys	2300
Db	5761	TGCTCTGTGTGTAGCTGTACCCCAAGGCTCTTGTCCAGAGCTGTGTACCTTGAGAT	5820	Db	6841	CCCGTGTAGATCAAGCTTCATGATGATGAGAGCTTCTGCCCCCGGCGCTGACAA	6900
Qy	1941	G1yG1nCySProCySAspProG1yVal1LeG1yArgG1nCySAspArgCySAspAspPro	1960	Qy	2301	ProVal1ThrValG1nPheArgLeuG1uThrG1uG1uArgTyrPThrPro1Le1CySVal	2320
Db	5821	GGCCAGGTGTCATGCAAGCAAGGTGTCTGGGGGTGATGATGATGATGATGATGATGAT	5880	Db	6901	CCCGTACGGTCAAGTCCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	6960
Qy	1961	PheAlaG1uVal1ThrThrAsnG1yCySG1uValAsnTYrAspSerCySProArgAla1Le	1980	Qy	2321	PheTyrAsnH1Ser1LeLeuValSerG1yThrG1yG1yTYTPTPTPTPTPTPTPTPTPT	2340
Db	5881	TTTGT	5940	Db	6961	TTCTGTGAACATTCATCTTGGTCAGTGGCACAGGAGCTGTGGTGGCGCAGAGGCTGAA	7020
Qy	1981	GLuAlaG1y1LeTPT	2000	Qy	2341	Val1ValPheArgAsnG1uSerH1sValSerCySG1nCySAsnH1sMetThrSerPheAla	2360
Db	5941	GAGGCTGGAGATCTGT	6000	Db	7021	GTCGTCTTCCGAAAG	7080
Qy	2001	LysG1ySerPheG1yThrAlaValArgH1sCySAspG1uH1sArgG1yTYTPTPTPTPT	2020	Qy	2361	ValLeuMetAspValSerArgArgG1uAsnG1yG1u1LeuProLeuTYrThrLeuThr	2380
Db	6001	AAAGGCTCTTTGGAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	6060	Db	7081	GTCGTCAATGACGTTTCTCGCGGAGAGATGGGAGATCTGCCACTGAGACACTGACA	7140
Qy	2021	AsnLeuPheAsnCySThrSer1LeThrPheSerG1uLeuLysG1yPheAlaG1uArgLeu	2040	Qy	2381	TyrValAlaLeuG1yVal1ThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu	2400
Db	6061	AACTCTTCAACTGACGTCATCATCTCTCAGAACTGAAAGGAGCTTGTGTGTGTGTGT	6120	Db	7141	TACGTGGCTCTAGGTGTCACTTGGCTGGCTTGTGTGTCACTTCTTCTTCTCACTTC	7200
Qy	2041	GLuArgAsnG1uSerG1yLeuAspSerG1yArgSerG1nG1nLeuAlaLeuLeuLeuArg	2060	Qy	2401	LeuArgG1LeuArgSerArgG1nH1sG1y1LeArgArgAsnLeuThrAlaAlaLeuG1y	2420
Db	6121	CAGCGAAATGATGACAGGCTTAGACTCAGAGGAGCTCCAGACAGCTAGGCTGTCTGCGC	6180	Db	7201	TGTGATATCTGTGCTCCAAACACAGGAGATCCGACCTGACACAGCTGCCCCGGGC	7260
Qy	2061	AsnAlaThrG1nH1sThrAlaG1yTYrPheG1ySerAspVal1LysValAlaTYrG1nLeu	2080	Qy	2421	LeuAlaG1nLeuValPheLeuLeuG1y1LeaG1nAlaAspLeuProPheAlaCySThr	2440
Db	6181	AAAGCCACGAGACACAGCTGTGCTACTTGTGCAAGCACTCAAGGTGGCTTACAGCTG	6240	Db	7261	CTGGCTACAGCTGTCTTCTCTCGGAGATCAACAGGCTGACCTCTTTTGTGCTGAC	7320
Qy	2081	AlaThrArgLeuLeuAlaH1sG1uSerThrG1nArgG1yPheG1yLeuSerAlaThrG1n	2100	Qy	2441	Val1LeaAla1LeuLeuH1sPheLeuTYrLeuCySThrPheSerTYrAlaLeuLeuG1u	2460
Db	6241	GCCACCCGCGTGTGTGCGCCAGAGACCCACAGCGGGCTTTGGGCTGTGTGCAACAG	6300	Db	7321	GTCATTTGCATCTGTGCTGACCTTCTGTACCTGTGACACTTTTCTGCGGCTGTGCGAG	7380
Qy	2101	AspValH1sPheThrG1uAsnLeuLeuArgValG1ySerAlaLeuLeuAspThrAlaAsn	2120	Qy	2461	AlaLeuH1sLeuTYrArgAlaLeuThrG1uValArgAspValAsnThrG1yProMetArg	2480
Db	6301	GAGGTGACTTCACTGAGATCTGTGCGGGTGGGAGCGCTCTCTGACACAGGCAAC	6360	Db	7381	GCTTGTGACCTGTATCCGGGACATCACTAGAGTGCAGATCAACAGCGGCCCATGGCC	7440
Qy	2121	LysArgH1sTYrG1uLeuLeu1LeG1nG1nThrG1uG1yG1yThrAlaTYrLeuLeuG1nH1s	2140	Qy	2481	PheTYTYrMetLeuG1yTYrG1yValProAlaPhe1LeThrG1yLeuAlaValG1yLeu	2500
Db	6361	AAAGCGCACTGGAGCTGATCCAGACAGACAGAGGTGGACCGCTGTGGCTGTCCAGCAG	6420	Db	7441	TTCTACTACATGCTGGGCTGGGCGCTGCTGCTTCAACAGGCTAGCCGTGGGCTG	7500
Qy	2141	TyrG1uAlaTYrAlaSerAlaLeuAlaG1nAsnMetArgH1sThrTYrLeuSerProPhe	2160	Qy	2501	AspProG1uG1yTYrG1yAsnProAspPheCySThrPheSer1LeTYrAspThrLeu1Le	2520
Db	6421	TAGGAGGCTTACCGCAGTGTGCGCCAGACATCGGCAACACTTAAAGCCCTTTC	6480	Db	7501	GAACCCGAGGGGTACGGGAAACCTGTACTGTGCTGCTCTCATCTAATGACAGCTCATC	7560
Qy	2161	ThrT1Val1ThrProAsn1LeVal1LeSerVal1ValArgLeuAspLysG1yAsnPheAla	2180	Qy	2521	TyrSerPheAlaG1yProValAlaPheAlaVal1SerMetSerValPheLeuTYr1Leu	2540
				Db	7561	TGGAGTTTGTGCGCCGGGTGCTTGTGCGTGTGATGAGTGTCTTCTGTACATCTGT	7620



NAME/KEY: CDS  
 LOCATION: (14989) .. (15130)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (15312) .. (15432)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (15856) .. (16039)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (16226) .. (16327)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (17669) .. (17763)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (17972) .. (18181)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (18702) .. (18885)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19001) .. (19129)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19231) .. (19411)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19545) .. (19712)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19815) .. (19941)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (20054) .. (20254)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (20547) .. (20721)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21048) .. (21160)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21338) .. (21417)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21507) .. (21633)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21721) .. (21843)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22399) .. (22526)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22761) .. (22849)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22954) .. (23148)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (23287) .. (23457)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (23557) .. (23791)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (24143) .. (24170)  
 US-10-292-798-931

Alignment Scores:  
 Pred. No.: 0  
 Score: 12339.00

Length: 24370  
 Matches: 2907

Percent Similarity: 36.924  
 Best Local Similarity: 36.924  
 Query Match: 79.384  
 DB: 12  
 Gaps: 32  
 US-09-916-849a-3 (1-2923) x US-10-292-798-931 (1-24370)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 DB 201 ATGGAGAGCCGGCCGACCGGGCTCCCTCCCAAGCCCGCGCGCTGCTGCTG 260

QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysatSerLeu 40  
 DB 261 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320

QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyThrLeuCysProSer 60  
 DB 321 GGATCCAGGAG 380

QY 61 SerAlaSerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 DB 381 TCAGGCTCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440

QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100  
 DB 441 GGCCACCTGGTACCCACACACATGCTGAGGGTTGGTGTTCAGAAATCCAGGCCCAT 500

QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
 DB 501 ATTCCTCTACACAGAGCTCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560

QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGluGluHisAspProCysLeuVala 140  
 DB 561 GGCCACCTTTCCCAAGAGGCAAGCTCACATGCTCCGAGAGAGACCCGCTTAAGGCT 620

QY 141 ProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
 DB 621 CCACGGCTCAGATGCAATGCTGCAAGCTGCAAGGCCCCCGGCTCAGGGAGGAGAA 680

QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgLysArgAsnValAsnThrAlaProGln 180  
 DB 681 AGCTACCAAGAAAGTCCCTGGGTGGGCTCGGAAAGAAATGTAATACAGCCCCCAG 740

QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 741 TTCCAGCCCCCAGCTACCAAGGCCACAGTCCGAGAAACACAGCAGCAGCCCTGTT 800

QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluTyrThrMet 220  
 DB 801 GCATCCCTGAGGCCATTCACACCCGACGAGGTGAGGCGAGTGCAGTGAATACACATG 860

QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 DB 861 GATGCCCTCTTTGATAGCCGCTCCACAGTCTCTTCCCTGAGCCCACTGATGCA 920

QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrLysSerThrHisValPheArgValThr 260  
 DB 921 GTAAACCAAGCCGAGAGAGCTGATGTGAGACCAAGAGCAACCAAGTCTTCAAGGGTCA 980

QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 981 GGCGAGGACGAGGAG 1040

QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGlu 300  
 DB 1041 GACACCAATGACCAATGACCTGCTGTTCCAGCAGCAGAGTACAGAGAGAGAGAGAG 1100

QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValaGlnAlaThrAspGlyAspAlaProPro 320  
 DB 1101 AACCTGAGGTGGCTATGAGGTGCTCACTGTCAGGGCCACGAGTGTATCCCTCC 1160

QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGluGlySerProSerGluValaPhe 340

Db 1161 AATGCCAATATCTGTACCGCTGCTGAGGGGCTCTGGGGGACGCCCTCTGAAATCTTT 1220  
Qy 341 GIUIIAspProArgSerGIYValIleArgThrArgGIYProValAspArgGIUVal 360  
Db 1221 GAGATCGACCTCGCTCTGGGGTGATCCGAACCCGCGGCTCTGTGATCGGAAAGGGTG 1280  
Qy 361 GIUSerTYRGIULeuThrValGIUValAspArgGIUValArgAspProGIYProArgSer 380  
Db 1281 GAATCCATCAACAGTACCGGTAGAGCGAATGACCAAGGTCGGGACCCGGGCTCTCGAGT 1340  
Qy 381 ThrThrAlaAlaValPheLeuSerValGIUAspAspAspAspAlaProGIUInPheSer 400  
Db 1341 ACCACAGCCGCTCTTTCTTTCTGTGAGAGATGACATGATATATGCCCCAGTTTATGT 1400  
Qy 401 GIUValArgTYRValValGIUValArgGIUAspValThrProGIYAlaProValLeuArg 420  
Db 1401 GAGAAAGCGCTATGTGTGTCAAGGTGAGAGGATGTGATCCAGGGGCCCAAGTATCCGA 1460  
Qy 421 ValThrAlaSerAspArgAspArgGIYSerAspAlaValIleIleTYRSerIleMetSer 440  
Db 1461 GTCAAGCCTCGAATCGAACAAGGGGAGCAATGCCGTGTGCTACTATAGCATATGAGT 1520  
Qy 441 GIYAspAlaArgGIYGIUInPheTYRLeuAspAlaGIUInThrGIYAlaLeuAspValIleSer 460  
Db 1521 GGCATATGCTCGGGGACAGTTTATCTGATGCCAGACTGAGAGCTCTGATGTGTGAGC 1580  
Qy 461 ProLeuAspTYRGIUInThrTYRLeuGIUInTYRLeuArgValArgAlaGIUAspGIYGIY 480  
Db 1581 CCTCTTGACTATGAGACGACCAAGAGATGACCTTACGGGTGCGAGCAAGAGTGTGTGC 1640  
Qy 481 ArgProProLeuSerAspValSerGIYLeuValIleThrValGIUValLeuAspIleAspAsp 500  
Db 1641 GGTCCCCCACTCTTAATGTCTGTGCTGTGAGACATGACAGTCTGTGATATCAACAC 1700  
Qy 501 AsnAlaProIlePheValSerThrProPheGIUInThrValLeuGIUInSerValProLeu 520  
Db 1701 AATGCCCATCTTGCTGACGACCCCTTCCAGGCTACTGCTCTGAGAGTGTCCCTTA 1760  
Qy 521 GIYTYRLeuValLeuIleValGIUValIleAspAlaAspAlaGIYAspAspAlaArgLeu 540  
Db 1761 GGCCTACCTGTCTCATATGCTGAGCTATCGACGCTGATGTGTGAGACATGCCGCTG 1820  
Qy 541 GIUValArgLeuAlaGIYValIleIleAspPheProPheThrIleAspAspGIYThrGIY 560  
Db 1821 GAATACCGCTGTGCTGGGGTGAGACATGCTTCCCTTCCATCAACATGAGCAAGC 1880  
Qy 561 TrpIleSerValAlaAlaGIULeuAspArgGIUValAspPheTYRSerPheGIYVal 580  
Db 1881 TGGATCTGTGTGCTGTGATCTGACCGGGAGAGTTGATTTCTACAGCTTTGGGTA 1940  
Qy 581 GIUValAspAspPheGIYThrProAlaLeuThrAlaSerAlaSerValIleThrVal 600  
Db 1941 GAAGCTCGAGCAATGCGACCTCCAGACATCACTGCTCGGCGAGGTGACGCTGACTG 2000  
Qy 601 LeuAspValAspAspAspAspProThrPheThrGIUInProGIUInTYRValArgLeuAsn 620  
Db 2001 CTGGATGTCACGACCAATCAACATCTTACCAACAGAGTACACAGTGGGCTCAT 2060  
Qy 621 GIUAspAlaAlaValGIYThrSerValIleThrValIleSerAlaValAspArgAspAlaHis 640  
Db 2061 GAGGATGAGCTGTGGGACCAAGCGTGTGACGCTGTGACCGTGTGACCGTGTCTCAT 2120  
Qy 641 SerValIleThrTYRGIUInIleThrSerGIYAsnThrArgAspAspPheSerIleThrSer 660  
Db 2121 AGTGTCACTACCTACCAATGACAGTGTGACATCTCGAAACCGCTTCTCCATCCACAGC 2180  
Qy 661 GIUSerGIYGIYLeuValSerLeuAlaLeuProLeuAspTYRLeuGIUInArgGIU 680  
Db 2181 CAAGGT 2240  
Qy 681 TYRValIleValAlaValIleThrAlaSerAspGIYThrArgGIUInAspThrAlaGIUInIleValVal 700  
Db 2241 TATGT 2300

Qy 701 AsnValIleThrAspAlaAsnThrIleArgProValPheGIUInSerSerIleTYRThrValAsn 720  
Db 2301 AATGTACCAAGCGCAACCAACCATGCTGTCTTTCAAGAGCTCCCATATACGTATAT 2360  
Qy 721 ValAsnGIUAspArgProAlaGIYThrThrValValIleIleSerAlaThrAspGIUAsp 740  
Db 2361 GTTATATAGGACCGGCGGAGGACCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2420  
Qy 741 ThrGIYGIUAspAlaGIYIleThrTYRPhMetGIUAspSerIleProGIUInPheArgIle 760  
Db 2421 ACAGTGAAGATGCCCGCACTACCTACTTATGAGAGACAGATCCCGCAAGTGTCT 2480  
Qy 761 AspAlaAspThrGIYAlaValIleThrThrGIUAlaGIULeuAspTYRGIUAspGIUValSer 780  
Db 2481 GATGTACACAGGGGGCTGTACCAACCAAGCTGAGCTGTGATGTGAAGACCAAGTGTCT 2540  
Qy 781 TYRThrIleValIleThrAlaArgAspAspGIYIleProGIUInIleSerAspThrThrTYR 800  
Db 2541 TACACCTGTGCTATCTGT 2600  
Qy 801 LeuGIUInIleLeuValAspAspValAspAspAspAlaProGIUInPheLeuArgAspSerTYR 820  
Db 2601 CTGAGATCTCTGTGTGAGACAGATGATGACATGACATGCCCTCAAGTCTCTGAGACTCTAC 2660  
Qy 821 GIUInIleSerValTYRGIUAspValProProPheThrSerValLeuGIUInIleSerAlaThr 840  
Db 2661 CAGGGCAGTGTCTATGAGATGTGTGCCACCTTCACTGACCTGTGAGATCTGAGCACT 2720  
Qy 841 AspArgAspSerGIYLeuAspGIYArgValPheTYRThrPheGIUInGIYIleAspAspGIY 860  
Db 2721 GATGTATTTCTGTGATTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2780  
Qy 861 AspGIYAspPheIleValGIUInSerThrSerGIYIleValArgThrLeuArgArgLeuAsp 880  
Db 2781 GAGGTGATCTTATTTGT 2840  
Qy 881 ArgGIUAsnValAlaGIUInTYRValLeuArgAlaTYRAlaValAspIleGIYMetProPro 900  
Db 2841 CGAGAGAACGT 2900  
Qy 901 AlaArgThrProMetGIUValIleThrValIleThrValLeuAspValAspAspAspProVal 920  
Db 2901 GCCCGCACTTATGAGATGACATGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2960  
Qy 921 PheGIUInAspGIUInPheAspValPheValGIUInIleAspSerProIleGIYLeuAlaVal 940  
Db 2961 TTTGAGCAGATGTGATGT 3020  
Qy 941 AlaArgValIleThrAlaThrAspProAspGIUInGIYThrAsnAlaGIUInIleMetTYRGIUInIle 960  
Db 3021 GCCCGGTGTACAGCTGACCTGCCAGTGAAGGACCAATGCCAGATTATGTATACCAATT 3080  
Qy 961 ValGIUInIleAsnIleProGIUValPheGIUInLeuAspIlePheSerGIYGIUInLeuThrAla 980  
Db 3081 GTGAGGGCAACATCTCTGAGAGTCTTCACTGTGACATCTTCTTCCGGGAGCTGACAGCC 3140  
Qy 981 LeuValAspLeuAspTYRGIUAspArgProGIUInTYRValIleValIleGIUInAlaThrSer 1000  
Db 3141 CTGTGTACCTTATGATCTGAGAGACCGGCTGAGTACGTCTGTGATCCAGGCAAGCTCA 3200  
Qy 1001 AlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAspAspPro 1020  
Db 1021 AATGT 3260  
Qy 1021 ProValIleGIYAsnPheGIUInIleLeuPheAspAspTYRValIleThrAsnArgSerSer 1040  
Db 3261 CCAAGT 3320  
Qy 1041 PheProGIYGIYAlaIleGIYArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
Db 3321 TTTCCCTGGGGGT 3380

QY 1061 ThrTyrSerPheGluArgIlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3381 ACTTACAGCTTTGACGGGGAAATGAATCAAGCTGGTCCGTCTCAATGCTCCACGGGT 3440  
 QY 1081 GluLeuIlySerAsnArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
 Db 3441 GAGCTGAGCTTAAGCCCGGCACTGGAACAACCGGCTCTGGAGGCCATCATGACGGTG 3500  
 QY 1101 LeuValSer----- 1103  
 Db 3501 CTGGTGTTC-AGGTAAAGAAAGGCCCAAGTGGCGCTGGGGGTGGGGTAACTCGCGGGATG 3559  
 QY 1103 ----- 1103  
 Db 3560 GTCTGGGCAACCACTGAGAGTGGGTGCGATCAGAAAGCAAGTACAGATCCACTCCCT 3619  
 QY 1103 ----- 1103  
 Db 3620 GCCCAGTGCTTGCAAGAGCAGAGAGGGGCAAGAGCAAGCTTGGAGAAAGCCCAAGAA 3679  
 QY 1103 ----- 1103  
 Db 3680 CCGTGGGGGCTGGTGGCAGAGCCCCACCTTCTGTGGGCTGGGCTGGGCTTCCGTTGGC 3739  
 QY 1103 ----- 1103  
 Db 3740 TGCCCCACTGTCTACAGACCGCTGACGTGGTGGGGAAAGTGTGTGAGGCTGCCCGGCG 3799  
 QY 1103 ----- 1103  
 Db 3800 TATGGCTGCCAAGATTGTGAAAAAGCGGTGAAAGCTGGGGGGCCAGAGAGCTGG 3859  
 QY 1103 ----- 1103  
 Db 3860 GGGAGAGGCAAGAGACGTGTGGCGCAGAGAGAGGCGAGGCCCTTCAACCC 3919  
 QY 1103 ----- 1103  
 Db 3920 CATCAAGATATTCTGACATCTACACAGCCTTGGCAGAGCTGGAGCCACGCGAGACAC 3979  
 QY 1103 ----- 1103  
 Db 3980 AGGGCTGCCGCAACCAAGGGGTGAGACCTGTAGTCTCTTCTGGCCCTTTTGCTCC 4039  
 QY 1103 ----- 1103  
 Db 4040 TGGCTGAGCTGCAGAAATCGGCTGACAGGACTTGTCTAGCGCATTTCTTCCCTCCAGG 4099  
 QY 1103 ----- 1103  
 Db 4100 CAGAAAGACCTGGCTGGCTCCCCCTTCAGAGCTCCCAAGAAAGTGAAGGCCAGAGCTC 4159  
 QY 1103 ----- 1103  
 Db 4160 ATTGTCTGTCCAAACAGACCCCACTTTCAGAGTCCGTGTTCGGGTCTGGCTTGAAGAG 4219  
 QY 1103 ----- 1103  
 Db 4220 GGGCTGCAATGAAGCTGGGGGGCAGGCTGAGGCTCGTGGGCCCCCTTGCCACTCATG 4279  
 QY 1103 ----- 1103  
 Db 4280 GGTGTCTTAAAGGTGCGCAGGCTTCTTCTCTAGCTGTGAACAGCCCAAGCTCAG 4339  
 QY 1103 ----- 1103  
 Db 4340 GAATGAGTGTAGGAGCAAGAGCAGGCGAGTGAAGTGTGAGCTTTGAGGCGAGAGACC 4399  
 QY 1103 ----- 1103  
 Db 4400 CAAGGACCTTATCTGAGAGAGAAAGAGTCCAGGGCGTGGAGCTGGCGGAAACC 4459  
 QY 1103 ----- 1103

Db 4460 AGGCTGGGCAATTCGTGCTTTGGCCAGCACTCAAGCTCAGACTCCTGGAACAGCAG 4519  
 QY 1103 ----- 1103  
 Db 4520 GGCTGGGGAGAGAGTGGAGGGGGTGGTGAAGATATTCTGTCTTTTATACAGCTG 4579  
 QY 1103 ----- 1103  
 Db 4580 TGGGAGCCAGAGGTAGCAGCAGGAGAAAGCAACCCAAATTGCTGAGTCCCTCC 4639  
 QY 1103 ----- 1103  
 Db 4640 AAGCTGGGAACAATGAGACCTCAATGAGGGATGCTTTAGCTGACCCAGGGGTGG 4699  
 QY 1103 ----- 1103  
 Db 4700 ATCAGGTCAATCCCTTAGCGACCTTCAATTCATCAGGAGGAGAAAGAAAGA 4759  
 QY 1103 ----- 1103  
 Db 4760 GGCTGGCAGGCCCTTACTCCACCCAGGAGTGTGGCTGAAGGGAAAGAGGGTATG 4819  
 QY 1103 ----- 1103  
 Db 4820 CAGAGTCAAGGGGCTGACCTCACTCAGGGTCTTGTGGAAGATGCTCCCGCCAC 4879  
 QY 1103 ----- 1103  
 Db 4880 CTTCAAGTAGCCACAGAAACAGAGAGAGCTTTCAAGGACTGTGCTGAGGTT 4939  
 QY 1103 ----- 1103  
 Db 4940 CTGAGCTCTGTGAGATTGATTTCCCTCCAGAGGCGGAAAGAGAGGCTGAGACTCC 4999  
 QY 1103 ----- 1103  
 Db 5000 TGTGTCTCAACTGTGCTGGGCACTTGAGACAAAGTGGAAAGAGTGGAAAGAGGGCC 5059  
 QY 1103 ----- 1103  
 Db 5060 TCTCCCTGTGTCTTCACACCCAGGCTGCTTGTGAGCGTGGCCAGGACTGGCAGA 5119  
 QY 1103 ----- 1103  
 Db 5120 GAAGCGCCCCAATTAAGCTTGGGTCTCCCAAGGCGAGAGGCTTAGACTTGTTC 5179  
 QY 1103 ----- 1103  
 Db 5180 TAGCTTAGAAGAGCTCGGGCAACATCCAGACGCGCTCCTCTGGCCCGGAGCTTGG 5239  
 QY 1103 ----- 1103  
 Db 5240 CCCACCTCTCCCTTTCGGGCTGACCTTGAAGGAGTCAACAGGCAACCCCAAGTT 5299  
 QY 1103 ----- 1103  
 Db 5300 CAGGAGCGGGGTGAGGGGAGCTGACTCCACTGCCCCCTTTGTTTCTCTCTTGT 5359  
 QY 1103 ----- 1103  
 Db 5360 TCTTGTCTCTCTTCTTCTTCCGCCCCGACATGAGAGCCCTGACAGACTGCTCTC 5419  
 QY 1103 ----- 1103  
 Db 5420 TCCGATCCCTCTCCCTGCTCTGCGGCTCTCGGCTTTCACACTTGCCATTTGCTGCT 5479  
 QY 1103 ----- 1103  
 Db 5480 CCTCTCTGAGGCTTGTCTTCCGTTTGTGATCCGATTTGTGCTCTCACTCTCTTC 5539  
 QY 1103 ----- 1103

Db 5540 CCTATGCTGCACTGCCCCAGCCTGGCCGATGCTCCCTGGAGGCTGAATTCCTC 5599  
Qy 1103 ----- 1103  
Db 5600 GCCTCTTCTTCTCTGTCTAGACACGACAGCTCCTGTCCATGTCTGGGACGCTCC 5659  
Qy 1103 ----- 1103  
Db 5660 AGAAGACAGAGCCCTCTTGCCCTCCAGTCTCCGCTGGACATCTGCCAGGGGAGAG 5719  
Qy 1103 ----- 1103  
Db 5720 GCTGGAGGCTGACCGGGAAGCTGGCACTTTCCAGTGTGCTGGCTAAGCTTGGAC 5779  
Qy 1103 ----- 1103  
Db 5780 AGGTGCAATTCCTGGGGTCAAGAGCCATCCAGCTGTCTTTCTGTCCAGAGAGCA 5839  
Qy 1103 ----- 1103  
Db 5840 GCTCAGGCAAGTACTCCCAACCCCTGTCCATCTGCTTTCACCTGAGGGTGACC 5899  
Qy 1103 ----- 1103  
Db 5900 TAGGTCAAGGCAAGGGAGGGTGATATAGAAAGAGAGCAGTGGTTGACTTGGGGA 5959  
Qy 1103 ----- 1103  
Db 5960 TGGGGGACACAGAGAGGGGGGAGTGGCCACAGATGTGGACAGATGGGAAGCA 6019  
Qy 1103 ----- 1103  
Db 6020 GAGCATGCTGGAGAGGGGCCAGGGGCTTAAGACATGGCGGTGGGGACACAGG 6079  
Qy 1103 ----- 1103  
Db 6080 GAGAAGTGTGTTCCCAAGGCTGAGGGGTGGCAAGTTTCCCGCTCATGGCTTGG 6139  
Qy 1103 ----- 1103  
Db 6140 TTGCCCCCTCTCTCTTATATAGTTATGGGCACTGTGTGTGAGTCTTGGCTGCTT 6199  
Qy 1103 ----- 1103  
Db 6200 CTGCTGAGAGGAATATCCGTCAAGGAATCATTAGGAGATATACAGGTATCAAGAG 6259  
Qy 1103 ----- 1103  
Db 6260 CTCTGTGCAAGTGAAGCTCTGTGGTGGGGGAAGTCTGAAACGAGAGAAAGTAGACC 6319  
Qy 1103 ----- 1103  
Db 6320 TGATGACATAAATCTGGGCAAGCTTCTGAGCAGGTGAGTAAAGGCAAGTTAAGACT 6379  
Qy 1103 ----- 1103  
Db 6380 CAGAGACTGTGGCTGTGTCTGGGTGGAGAGACAGCTGTGACAGAGCTTTGAGAC 6439  
Qy 1103 ----- 1103  
Db 6440 GAGAGCAGAGAGAACAGAAAGTGGGTATGACAGAGGCTTGTGTGAGAGCTTAGAG 6499  
Qy 1103 ----- 1103  
Db 6500 AATAAGCAATGATAGGCTGAGGGTAGTGGGCAGTGGGTGAGACATGGCTGAGAT 6559  
Qy 1103 ----- 1103  
Db 6560 GGGGCTGTCTTTAGCCTTCCGAGCTTCAAGTACCAACGCTGTATCACTCAAGCA 6619  
Qy 1103 ----- 1103  
Db 6620 GCAAGTAGTACACACAGATGCAGAATTGTCCCATGTCCCTTGGCTACTCCACCCC 6679

Qy 1103 ----- 1103  
Db 6680 ACCCAGAAAACATGGGCTCTGTGCCGTTACTGTAGGGTAAGCATTTGTTTATATC 6739  
Qy 1103 ----- 1103  
Db 6740 TAGTGGGCTCCCTGTATGTGTGTTGTGTTGTGTTGGAGACAGTCTCACTTG 6799  
Qy 1103 ----- 1103  
Db 6800 TCAACCAAGCTGAGTACATGGCCGATCTGGCTCATCTGCAACCTTGTCTGGGT 6859  
Qy 1103 ----- 1103  
Db 6860 TCAAGCAATTCCTGTCTCAAGCTCTGATATCTGGAGCTCTGAGTAGCTGAGATT 6919  
Qy 1103 ----- 1103  
Db 6920 ACAG3ACCCACCAACGACCCCACTAATTTTGTATTTTATAGAGACGGGGTTTAC 6979  
Qy 1103 ----- 1103  
Db 6980 CATGTGTGATGATCTGATCTGAATCTGAACTCTGATCCACCCGCTCAAGCTCCA 7039  
Qy 1103 ----- 1103  
Db 7040 AAGTTGGGACTACAGGGGTGAGCCATGGCCCGGCCCTGTATGTTTAAAGTCC 7099  
Qy 1103 ----- 1103  
Db 7100 GTTATTCCTTGTGAGACTGATATAGATTGTATCCACATCTCAGTGTCTGAACACC 7159  
Qy 1103 ----- 1103  
Db 7160 TACTATGTGAAGCATTTGTAATTAATACAAATATGATATGCAATGATGCTGTCTGT 7219  
Qy 1103 ----- 1103  
Db 7220 AATCCAGTCTTAGAAGCTGAGATGAGATCACTTGAAGCCAGAAATTTAGACCA 7279  
Qy 1103 ----- 1103  
Db 7280 GCCTGACACACAGAGAGCCCATCTTACACAAACAAATTTAGCAGCATAGT 7339  
Qy 1103 ----- 1103  
Db 7340 TGTAGCACTGTGTCTAGCTAATTTGGAGGTTGAGTGGAGAGATGCTTGAAGCCA 7399  
Qy 1103 ----- 1103  
Db 7400 GAGTTGAGGCTACAGTAGATATATGTGCTCCAGTGTAGGCGACAGAG 7459  
Qy 1103 ----- 1103  
Db 7460 AGTCTGTCTTAAAAAAGAAAAAAGACTATCCCTTCTTGGCTGTAAAGCAATTT 7519  
Qy 1103 ----- 1103  
Db 7520 CTTCTTCTCAGGCCCCAGCTTTCATTTTATAGCTAATTAATTAATTCATTCATGTGTA 7579  
Qy 1103 ----- 1103  
Db 7580 TGTGATTCAGGTGTATAGTATTTCTGGCTTAGGGTGTGTGGATGAGACACC 7639  
Qy 1103 ----- 1103  
Db 7640 AGCAAGTGTATTAAGTCTTAAGCTTTATAGTAGTGTGTGCTGGGGCATAGCA 7699  
Qy 1103 ----- 1103  
Db 7700 GGCATTTGGGTATATGTGTTAGCCAAAGTGGATGATGATGACTAGAGGCTTAGACCA 7759



QY 1103 ----- 1103  
Db 7760 GAGAGAGACTGACTGAGTGGGGCTGTGGGCTCTCAGAGTGAAGCTGCCAGGCCCTT 7819  
QY 1103 ----- 1103  
Db 7820 CTATCTCTGCTTCCGGTCCCCCAACCCGACTCAGCTCTGTCTCCAGCATGGGGCAGTGGG 7879  
QY 1103 ----- 1103  
Db 7880 TGATGATGTCACTGGTTAACGTGGCTCAGTGAATTCCTGGGGTGGGGCCCCCAGGCT 7939  
QY 1103 ----- 1103  
Db 7940 ACCCAAGACGCCAGAGAGTGGGGGAGAAATGCTTTGTGTGATGTTCATGGG 7999  
QY 1103 ----- 1103  
Db 8000 GATAGCAGTCTGGCGCCGCAAGCAGCCATGGGGCTCAGTGGAGGGGAACTAGGTGCC 8059  
QY 1103 ----- 1103  
Db 8060 CACTGACCCAGTTTACCACTCCCCCACCAGCCCACTCCGTGTGTGTGGGGAG 8119  
QY 1103 ----- 1103  
Db 8120 TGGGAAACAGACCATGTGTGCATTAATGTGTGTGAGGTGTAGAGGCTGAGGGGCA 8179  
QY 1103 ----- 1103  
Db 8180 GGGGATGATTTTGTGTGTCTCAGATGTCTGTGGCCCTGTAGAAAATGTGAAAAGA 8239  
QY 1103 ----- 1103  
Db 8240 CAACATAGGTTCCCGGAGAGGAGCAGACAGCTGTGGGGAGAGGAGAGCGGTGGC 8299  
QY 1103 ----- 1103  
Db 8300 TTGGGAGGAGAGGTGACAAACATTAAAGGTGTCCGTGATCAGCGACAGAGGTT 8359  
QY 1103 ----- 1103  
Db 8360 TTGGGAAAGATGCAAAAGAACTGCAATCTCATCCCTGAGCCCTGAGAGAGTATGG 8419  
QY 1103 ----- 1103  
Db 8420 GCCCAGCTGACATACCTGGCTCAGCACCTTATGCAATTGGAAGAGAGCCTTCT 8479  
QY 1103 ----- 1103  
Db 8480 TGGTGGGTGTGAGAAATGAGGCTGAATGCAAGCCCGAGGCTGGCTCTGACTGTG 8539  
QY 1104 ----- 1119  
Db 8540 TCCCTCTCCACAGCGCGGTACACAGCGTGAACCCCAAGTGCCTGCGCTGTGACATC 8599  
QY 1120 TLeThrApGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Db 8600 ATCAACGATGAGTGTCAACCAAGCATCAAGCTGGCGCTGAGAGCAATGTACCCGAG 8659  
QY 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaAlaThrLeuAlaThr 1159  
Db 8660 CGCTTCCTGTACCACTGTAGGCTCTTCAACGAGCGGTGGCGGCAAGCTGGCCACG 8719  
QY 1160 ProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
Db 8720 CCACCGGACCACTGTGTGTCTTCAAGTACAGGGGACACCGACGCCCGGGGGGAC 8779  
QY 1180 TLeuLeuAsnValSerLeuSerValGlyGlnProProGlyGlyGlyProProPhe 1199  
Db 8780 ATCTCAACGTCGAGCTGTGGTGGGAGCGCCAGGGCCCGGGGGCGGCGCCCTTC 8839  
QY 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrlleuAsnArgSerLeuLeuThrAlaIle 1219

Db 8840 CTGCCCTCTAGAGACTTGAGAGAGCGCTTATACCTCAACCGCAGCCCTGTGACGGCCATC 8899  
QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGlyAsn 1239  
Db 8900 TCGGACAGCGCGTGTCTGCCCTTGACAGACAACTGTCTGGGGAGCCCTGGAGAAC 8959  
QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 8960 TACATGGCTGGCGGTGTGGTGTGGCTGTGACTTCCTCCGCGCTTCATCGCTCTCC 9019  
QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
Db 9020 TCGGTCTCTTCCGCCCATCAACCCGTGAGGGGTGGCGCTGCCGCCGCCGCCGCC 9079  
QY 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 9080 TTCAAGGTGACTACTGCGACAGCCAGGTGACCTGTCTACTGCGGCTGTGTGGCCC 9139  
QY 1300 HisGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 9140 CAGGGGCTGTGCGCAGCGCGGAGGGGGCTACACCTGCTGTGTGTATGGCTACAG 9199  
QY 1319 ----- 1319  
Db 9200 GGTGAGCCAGAGGAGGAGACTCATGGGCGAGCCCTGGAAGCTGAGTGTGAGAGC 9259  
QY 1319 ----- 1319  
Db 9260 ACAATTCAGGACAAATGTGGCGGCTGCTCATTTCTTCCGAGTGAAGTGCACCTCA 9319  
QY 1319 ----- 1319  
Db 9320 CTGAGAGTCAATAATCCGGCTGTCTCCTGACAGACCCCACTGAAGAAAGCATGCA 9379  
QY 1319 ----- 1319  
Db 9380 GAGGTCTTCTCCCTGTGTGACAGAGACACAGGCAGTGCATTGTGGCAGAGCTGG 9439  
QY 1319 ----- 1319  
Db 9440 CTCAGAGCAGTGGCGCTGTGTGTATACGCTTGGCCCTAGTAGCTGTGTGACT 9499  
QY 1319 ----- 1319  
Db 9500 TGAACAATTACTTAACCTTCTGCGCTCGGTTATTGAATGAGATATGACCGTAC 9559  
QY 1319 ----- 1319  
Db 9560 TTAGTCATAGTGTGTGTGAGATTAAATTAATTATTTGATCTTGAACACAGCTGG 9619  
QY 1319 ----- 1319  
Db 9620 TGGCATGGAAGTGTGGCTGTGATTTAAGGAACGCTGAGATTGGGTGACTTGTCTCG 9679  
QY 1319 ----- 1319  
Db 9680 GTCTCATCACACCGGTTTAGAGTGAAGCTGGGCTTGAACCTGTCTCCCACTG 9739  
QY 1319 ----- 1319  
Db 9740 AGGAGTCTCTTTCTGAGCTGACAGAAATCTTTTCTCCAGGATTAAAGGGAAGAC 9799  
QY 1319 ----- 1319  
Db 9800 ACAGAAAGCTGGGGAACTGAGAGGGGTTGAGTTCTGAGATGGGGTGTCCAGAAACC 9859  
QY 1319 ----- 1319  
Db 9860 ATTGACCTGTATGGGCTGGGAGAAATGCTCCGCTCCGCCAGCTGTGTAAAGA 9919  
QY 1319 ----- 1319

Db 9920 GGATCTGGCTAAGCGGCTTCCTGTGGCCTGTCAAGCCAAAGTCTTTACACATCTGTCT 9979  
 Qy 1319 ----- 1319  
 Db 9980 CCGTGGAGTGTCACTGTCTCTGTCTATAGCTCCTGAACCCCTCCCTTCAAGGCTCTGTGT 10039  
 Qy 1319 ----- 1319  
 Db 10040 CTTCCTCAGAACACCCCTGTCTCCCTTCCCTGTGGGCTTTCTTCCTAGATGCCAGGC 10099  
 Qy 1319 ----- 1319  
 Db 10100 CTTGCTCTGTCTCTCAAGGCGCATGCCCATCTGCTCTGCCAGACCATAGAGAGGCC 10159  
 Qy 1319 ----- 1319  
 Db 10160 CACTCTTCCCCGCTGCATCTTCCTGCCAAGACTTGCAGACTTGTCTCTGTCTG 10219  
 Qy 1319 ----- 1319  
 Db 10220 CTGCTAAGGGGTGGGGGAGAGGAGGACGGGCTTGGGGGAGGGAGAGAAGTCTTGTCTC 10279  
 Qy 1319 ----- 1319  
 Db 10280 CAGAGAGATGTGGGGGGGTTGGAGGGGGGGTTAGACAGCGCAACAAGTTTCC 10339  
 Qy 1319 ----- 1319  
 Db 10340 ATGGAACAAGAGCGCATCTAGGCAAGAGAGCTGAGAACTATCTTAGAATCTGGGA 10399  
 Qy 1319 ----- 1319  
 Db 10400 ATGGGAGAGAGAGAGAGGTGATTTTCCCATCTGATGTGGATTTTCAGCTCTCTGT 10459  
 Qy 1319 ----- 1319  
 Db 10460 TGAATCTTGTCTGTCTCAAGGCTCAGTTTCCCATCTCAGCAATGCCAGGTGCACAG 10519  
 Qy 1319 ----- 1319  
 Db 10520 AGGTCTTATGTCTGACCGCAGGGCTCAAGGTGGAGTCTCTGCCCATTTCTGGGGGC 10579  
 Qy 1319 ----- 1319  
 Db 10580 TCCCCAGCTTCTCTCTATACTTTCATTAACCCAGCAAAATCTAAGTGAAGCTGTCTC 10639  
 Qy 1319 ----- 1319  
 Db 10640 CCATCCCCAGGAGGAAATGATCTCACTGCTGAGAGGGAAGAAAGCACAGTGGGAA 10699  
 Qy 1319 ----- 1319  
 Db 10700 GAGAGAGAGAGAGGCTCCAGGCTCCGCTGTCTGACAGGCCATCTGTGTGTGCAC 10759  
 Qy 1319 ----- 1319  
 Db 10760 ATTATCCCTCTTGGGGGCGAGAGACCCCCAGGTGTGTGTGTGATGATCAGTGAAGGC 10819  
 Qy 1319 ----- 1319  
 Db 10820 TCCCGGATCCGGGCTTTACACCAAGTGTCCAGCAGGCTGTGTATGAAACCCAAA 10879  
 Qy 1319 ----- 1319  
 Db 10880 TGGGGTAGGAAAGTGTGTGGGACTCCAGAGAACTTGACAGGAGAGAGTGTGTGAGCTG 10939  
 Qy 1319 ----- 1319  
 Db 10940 TGGGAGAGCGGCTCAGTGAACCAACAAGGTGCCATGAAGGCTGGGGAACATAGGTC 10999  
 Qy 1319 ----- 1319  
 Db 11000 TCAGAGAAATGGACACAGAGAGAAAGTGGGATGCTCTGGAGGTTTGGGAGTATTAAG 11059

Qy 1319 ----- 1319  
 Db 11060 AGGATCAGGAGAGCCATAGGAGCTATCACTGGGTTGTGGGGGTCTCAGTTCCCTCCT 11119  
 Qy 1320 ----- 1325  
 Db 11120 GTCCAGGATCTAGTACTACCTGCTCTTCTCTGTTCAGAGTGAACACTGTGAGGTG 11179  
 Qy 1326 SerAlaArgSerGlyArgCysThrProGlyValCysLeuAspGlyGlyThrCysValAsn 1345  
 Db 11180 AGTGTGCTCAGAGCGGTGACCCGGGTGTCTGAAGAAATGGGGGACCTGTGTCAAC 11239  
 Qy 1346 LeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPheGlyLeuValGln 1365  
 Db 11240 CTGCTGTGGGGCGGTTTCAAGTGCATTCCTCCATCTGAGAGACTTGAGAAAGCCCTACTGC 11299  
 Db 11300 CAGGTACACAGCGCAGCTTCCCGCCCACTCCTCATCACCTTTCGGCGCTGCGCCAG 11359  
 Qy 1386 ArgPheHisPheThrLeuAlaLeuSer ----- 1394  
 Db 11360 CGTTTCACTTCACTGAGCCCTCTC-GTGAAGTGTGGGCACTGGGGGTGGGAGTGGG 11418  
 Qy 1394 ----- 1394  
 Db 11419 CCTGTGGGCACTGAGGTGCTCTCTGTGTGTGAAGTAAAGATCAGGCAGTGA 11478  
 Qy 1394 ----- 1394  
 Db 11479 AGCGTGACAGAGCTGCCGAGAGCCAGGTCTGTCTTGAGATTGCTGTGTCTGCC 11538  
 Qy 1394 ----- 1394  
 Db 11539 AGCAGATCTCCCTCCCTGCGCTGTGTAGGTGCGGGGTCTGGGATCAGGTGGGTAC 11598  
 Qy 1395 ----- PheAlaThrLysGlyLysArgAspGly 1402  
 Db 11599 CCCATTCCTGCCCCCATCCCAACTCTGTTCAGGTTTGCCACAAAGAGCGCGAGCGG 11658  
 Qy 1403 LeuLeuLeuTyrAsnGlyArgPheAsnGlyLysPheValAlaLeuGlyValIle 1422  
 Db 11659 TTGCTGTGTACAAATGGGGGTTTCAATGAGAACATGACTTGTGTGCTCTGAGGTGATC 11718  
 Qy 1423 GlnGluGlnValGlnLeuThrPheSerAla ----- 1432  
 Db 11719 CAGAGCAGGTCTCAGCTTCACTTCTGTGC-AGGTATCACAGTGTGCCCATCTTGGCC 11777  
 Qy 1432 ----- 1432  
 Db 11778 CATCTTCAAAAGGCCCAAGTCTTTCAGGCCCTGACCCCAAGCACATACTATCAGGC 11837  
 Qy 1432 ----- 1432  
 Db 11838 AAATCTGGGCCAGCCAGCCAGCACTGGCACCCCAACCTGTGCATCTTCAAGCCCTGTCA 11897  
 Qy 1433 ----- GlyLysThrThrValSerProPhe 1442  
 Db 11898 GGCATTCAGCTCACCTGATCTTCCCAAGGAGATCAACACAGGTGTCCCATATTC 11957  
 Qy 1443 ValProGlyGlyValSerAspGlyGlnTyrPheThrValGlnLeuLysTyrTyrAsnLys 1462  
 Db 11958 GTGCCCGGAGAGTCAGTATGCCAGGTGCATACGCTGACGTGAATAATCAATAAAG 12017  
 Qy 1462 ----- 1462  
 Db 12018 GTGGGTGGAGGGACAGAGAGGTGGGGGTTCTGTCTTTCCTCAAGTCTTACCAAC 12077  
 Qy 1462 ----- 1462  
 Db 12078 CTTAGTGGCAATTGCTCAGGCTTGGGTGGGTGTGAGGCAATTCTGGCTGAGAGGA 12137

[illegible]

Db	13218	AGAGCCCAAGCTCAGGTCCAGCTCAGCTGCTTCTCTCCATCTCCAGGCTGCCCTGCCAATA	13277
OY	1575	ysaenValCybaapSerenth-CyahiBaenGlyGlyThrCyVaIaenGlnTrpAspA	1595
Db	13278	AGAACTGTGTGTGACAGCAACATCTGGCCACAAATGGGGGCACTTGGCGAACAAGTGGAGCG	13337
OY	1595	laPheSerCyBgluCyProLeuGlyPheGlyGlyIysSerCyBaIaGln-----	1611
Db	13338	CGTTCACGTCCGAGTGTCCCTTGTGGGCTTTGGGGGCAAGACTCGCCCAAGGTTAGAGGG	13397
OY	1611	-----	1611
Db	13398	GCAGCTGTAGAAGGCCACAGCCTGGGTGCCATCAACAGTCTGGGAATCTGGAGGTTGGG	13457
OY	1611	-----	1611
Db	13458	GCAGGCACTGGGCGAGGCTCTGCTGAGCGGGGCTGTGGGTGAAAAAGCTGTCTGGGCAAG	13517
OY	1611	-----	1611
Db	13518	AGCCCTGAAGAGGGCAACACAGAGACAAGAGGGGCGAGTCAAGAGGCAAGGCTGGGCACAGC	13577
OY	1611	-----	1611
Db	13578	GTGGGGCGGTCCAGGGCAGGTACGCACTTGGAGAGGGCGGGGCTGATGAGGGAGTGGGC	13637
OY	1611	-----	1611
Db	13638	TCGTCTCCCGTCAGCCGCCACCGCTGAGCATCACGCCCAAGGGGCTCTGAGGCTCC	13697
OY	1612	-----GluMetAlaBspProGlnHisPheLeu	1620
Db	13698	ACCGGTACAGTCTGCCTTTCTGCGCTCCCGAAGAAATGGCCAAATCACAGCACTTCTCG	13757
OY	1621	GlySerSerLeuValAlaTrpHisGlyIleuSerLeuProIleSerGlnProTrpTrpLeu	1640
Db	13758	GGACGAGCCTGGTGTGTGGCAAGGAGCTCTGTGGCCCATCTCCCAACCCCTGTGATCTTC	13817
OY	1641	SerIleuMetPheArGThrArgGlnAlaBspGlyValLeuLeuGlnAlaIleThrArgGly	1660
Db	13818	AGCCTATGTTTCGGCAAGCGGCCAGCGCAAGGTCCTGTGTGAGGCCATCACAGGGGG	13877
OY	1661	ArgSerThrIleThrIleu-----	1666
Db	13878	CGACGACCATCACCTTACAGGTGATGATGAAAGGGCGGTGGCCCTGGGCTGGCCATA	13937
OY	1666	-----	1666
Db	13938	GGGCCCTGGTAGCCTTAGGGCGGCTGGACAAGAAATGGCTGGAGGTCTTGGGCAAGGCT	13997
OY	1666	-----	1666
Db	13998	GGGACATATAGAGCGCGCTGATCCGTTGGAAAGTCAATGCTGACCTGTGGGCT	14057
OY	1666	-----	1666
Db	14058	GAGGAAATTAATAGCTTTCTCTATAGCCTCAGCTGGGCAAGGGAGTGGTTGAT	14117
OY	1666	-----	1666
Db	14118	GGTTCAGAAAGGGGAACGTGAGCTGGGAACAAGTGAATTAAGCGGTCTGTAACCTGT	14177
OY	1667	-----GlnLeuArgGlyGlyHisIleValMetLeuSerValGluGlyThr	1680
Db	14178	TTTCTTCTCTCTGATGTGAGCTACAGAGGGGCAAGTATATCTGAAGCTGAGGGGCA	14237
OY	1681	GlyIleuGlnAlaSerSerLeuArgLeuGlyProGlyArgAlaAspHisAspTrpHis	1700
Db	14238	GGGCTTCAAGCCCTCTCTCTCCGTCTGAGGCCAGGCCGGGCCAAATGAACGTGACTGGCAC	14297
OY	1701	HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp	1720

[illegible][illegible]

QY 2004 ----- 2004  
Db 16517 GACTGTTTGTGTTTGTGTTTGTGTTTCTGGAGACAGTCTGTCACCC 16576  
QY 2004 ----- 2004  
Db 16577 CAGGCTGTGTGAGTGGCGGTGATCTCGCTGCTGCAACTCTGCTCCAGGTTCAAG 16636  
QY 2004 ----- 2004  
Db 16637 CAATTTTCTGCGCTAAGCTCTGGCGCGCACACCGGTGCTAATTTTTTTTTTTTTT 16696  
QY 2004 ----- 2004  
Db 16697 TTTTGTGATTTTGTAGAGACAGGTTTTCACATGTTGGCAGTCTGTCTGAA 16756  
QY 2004 ----- 2004  
Db 16757 CTCTGACCTCATGATCCGCTGCTGCGCTCCCAAAGTGTGGGATTAACAGCGTGAAG 16816  
QY 2004 ----- 2004  
Db 16817 CCACTGACCCCAACACATTTTGGACTTTTTTTTTTTTTTGTAGAGAGAGTCTGCTC 16876  
QY 2004 ----- 2004  
Db 16877 TGTTCATCAGGCTACAAATGACAGTGGCATGATCTCGCTCTGCAACTCACAAGG 16936  
QY 2004 ----- 2004  
Db 16937 TTCAAGGATTTCTCTACTCTGACGCTTCCAGATGCTGTGATCAAGCGCTGTGATCA 16996  
QY 2004 ----- 2004  
Db 16997 TGCCAGCTAATTTTTATATATTTTTTATAGAGACAGGTTTTCACATGTTGGCCAGA 17056  
QY 2004 ----- 2004  
Db 17057 TGGTCTCGATCTTTGAACCTGTGATGACGCCGCTGCGCTCCCAAAGTCCGAGATTA 17116  
QY 2004 ----- 2004  
Db 17117 CAGCGGTGACCAACCGTGGCCAGCCACATCTTGGACTTTTAAACAAATGCTGAAAAAC 17176  
QY 2004 ----- 2004  
Db 17177 CTGCACTGTGCTTGGCATAGGGGCTGTGTAAGATTCAAAAGATGATCATGACGGTTCC 17236  
QY 2004 ----- 2004  
Db 17237 AATTCGACGTCACACCGAGTCTGTGATTAAGCCACCAAGGACAGACTGTGGGG 17296  
QY 2004 ----- 2004  
Db 17297 AGTAGGGGGTGTGGCTGTGACCCAGAGACACCAAGCCAGATGGGGGCTGGCGCTGGT 17356  
QY 2004 ----- 2004  
Db 17357 CTGGAATGGAATGAGCAGGTTTCATAGGAGAGAGTGGGAGAGACAAAGTGTGGC 17416  
QY 2004 ----- 2004  
Db 17417 TGAAGACAGGACAGATGGGGGCTCAAGAGATGATGACCTGTTCCTTAGAGCCATG 17476  
QY 2004 ----- 2004  
Db 17477 TGTGACATGATCTTGGCGAGATTTTGGCCCTGTGTCTTACATACCTGTCTCTGTTT 17536  
QY 2004 ----- 2004  
Db 17537 CACGAGGCCCTGTGTGTAGTTACTGTCCCGGCCCATGAGGTGTGTGCTCCAGCAG 17596  
QY 2004 ----- 2004

Db 17597 AACCTGACAGCAGTCTCAGGGGCTTCTGTGTGTCTCCCTGAGGCGGCGCTTTGGCT 17656  
QY 2005 -----  
Db 17657 TCTTCCCCCAGGAGCTGTGTGGCGCTGATGATGACAGAGGGGTGGCTCCCCCAA 17716  
QY 2021 snleuPheanCystrSer1LeThrPheSer1leuLeuysglYpHe----- 2036  
Db 17717 ACCTTTCACTGCAAGTCAATCACTTCTTCAGAACTGAAGGCTTCGTAGTGAACCC 17776  
QY 2036 ----- 2036  
Db 17777 CTGATCTCAATCTTTTCCCTGTCTTGTCTGTGAGTCACTTGCCCACTCCATCTT 17836  
QY 2036 ----- 2036  
Db 17837 TGAAGACGGGCTTCTGAATTCAGCCTGTGTGTCTGTGGGCTCCAGCTGAGAGGCC 17896  
QY 2036 ----- 2036  
Db 17897 GTCTCACTCTGAGTCTTTTGTCTCAGAGTTCTGCTCACTCTGCTCCG 17956  
QY 2037 -----  
Db 17957 CTCCGTCTGTCTTGGCTGAGCGGCTTACAGCGGATGATGACAGGCTTACAGGCGCG 18016  
QY 2052 SerGlnGlnleuAlaIleuIleuArgAsnAlaThrGlnHisPheThrGln 2071  
Db 18017 TCCAGACGTAAGCCCTGTCTGCGCCAGCCACGACACACACACACACCTGCTTCTCGGC 18076  
QY 2072 SerAspValIysValAlaTyrgIleuAlaThrArgIleuAlaHisGlnSerThrGln 2091  
Db 18077 AGCGACGTACAGGTGCTTACCAAGCTGGCACGCGGCTGTGCGCCACAGAGCACCCAG 18136  
QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGln----- 2106  
Db 18137 CGGGGCTTGGGTGTGTCTGCCACACAGAGAGTGCATCTGAGGTGGGGCTTGGAGGA 18196  
QY 2106 ----- 2106  
Db 18197 TGAAGGCTGGCTGTGATAGGTGATGAGGTGATGAGCTGCTCATGGCAACTGGG 18256  
QY 2106 ----- 2106  
Db 18257 GGGCAGAGGGGCGCTCCATCCACCTTACAGAGACTCCTGTGTGGCAGAGGCTCCA 18316  
QY 2106 ----- 2106  
Db 18317 TGAACCTGTGACCCCTGGCCAGCCCTTCCCAACCTTCTCATATGTAATATGTG 18376  
QY 2106 ----- 2106  
Db 18377 TGTGGGGATGTGTGACCCAGAGACAGGCGCTGTGTGTGATCAATGGGCTGGAGGA 18436  
QY 2106 ----- 2106  
Db 18437 AGCATATCTCAGACATGTGTGTGGGTCCAGGCAAGGGCTGGAAAGCTTATATAGA 18496  
QY 2106 ----- 2106  
Db 18497 GAATGAGAGAGGGCTTAAAGGGCAGAGATCACTGACCTGAGTGGGACGTGCGGTCCA 18556  
QY 2106 ----- 2106  
Db 18557 GGAACATTTCAAGGGAGGCAACGTGAGCACAGAGGAGGGGCTGTGTGGGAGCAGGGA 18616  
QY 2106 ----- 2106  
Db 18617 CTGCCCCGGGCAAGGCGGGGCTCCATGTGTGGGATGTCAGGTGTGTGTGAGCACA 18676  
QY 2107 -----  
AsnIleuIleuArgValGlySerAlaIleuIleuAspPh 2118

Db 18677 CCCACCGTGAACCTTGCCCAACCCCAAGAAATCTGCTGGGGTGGGACGGCCCTCTGACAC 18736  
 Qy 2118 rAlaAsnLysArgHisThrProGluLeuIleGlnGlnThrGluGlyGlyThrAlaTrpLeu 2138  
 Db 18737 AGCCAAACAAGGAGGACCTGGGAGCTGATCCAGACACAGAGGGTGGCACCGCTGGCTGCT 18796  
 Qy 2138 uGlnHisThrGluLysAlaTrpAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSe 2158  
 Db 18797 CCAAGACATAAGAGGCTTAACGCAAGCTGGCCCTGGCCAGAACTGGGGACACTTACTTAAG 18856  
 Qy 2158 rProPheThrIleValThrProAsnIle----- 2167  
 Db 18857 CCCCCATCACTGCTCAAGCCCAACATTTGGTAAGGCTGGTGCCTGGGTTGGGAGGGGTT 18916  
 Qy 2167 ----- 2167  
 Db 18917 TGTGAGGAGAGTCCCGCAAGAGCGGCTGTGCGGGTCTGCTGCTCAAGTCCG 18976  
 Qy 2168 -----ValIleSerValValArgLeuAspLysGlyAsnPheA 2180  
 Db 18977 ATCTGTGACCATCCCTCTTCTTAAGTCATCTCCGTAGTGCCTTGACAAAGGGAACTTTG 19036  
 Qy 2180 lAsGlyAlaLysLeuProArgTrpArgLysAlaLeuArgGlyGluGlnProProAspLeuGluT 2200  
 Db 19037 CTGGGGCCAAAGCTGCCCGCTACAGAGCCCTGCGTGGGAGACAGCCCGGAGCTTGAAGA 19096  
 Qy 2200 hTrpValIleLeuProGluSerValPheArg----- 2210  
 Db 19097 CAACAGTCATCTCTCGTAGAGTCTGTCTTCAG-AGGTCAGTGTGGCCATGGAATTGAGTTG 19155  
 Qy 2210 ----- 2210  
 Db 19156 GGAGCTGGAACCCAGTGTCTGTGAGACTCCACAGAGAGAGGCGCCAGCTTAAGTGTGAC 19215  
 Qy 2211 -----GluThrProProValValArgProAlaGlyProGlyGluAlaGlnG 2226  
 Db 19216 AGTGTCCCTCCCAAGAGACGCCCCCTGTGTCAAGCCCGACAGGCCCGAGAGAGCCCAAG 19275  
 Qy 2226 lProGluGluLeuAlaArgArgLysArgHisThrProGluLeuSerGlnGlyGluAlaVal 2246  
 Db 19276 AGCCAGAGAGAGTGGCAAGGCGACAGCGCACCGGAGCTGAGCCAGGAGTGAAGCTG 19335  
 Qy 2246 aAlaAsnValIleIleThrArgThrLeuAlaGlyLeuLeuProHisAsnTyrTrpProA 2266  
 Db 19336 TGGCCAGGTCATCATCTTACCGCACCTCGCGCGGCTAATGCTCATTAATGACCTTG 19395  
 Qy 2266 sPlyAsnSerLeu----- 2270  
 Db 19396 ACAAGCCGACGTTAAGTTCAGACAGTGAAGGGAACAGTGTGGGTATGGGTCCGGCC 19455  
 Qy 2270 ----- 2270  
 Db 19456 GGTGAGTGTGAGGACATGAGAGGGGTGCGGGGCGTCTCCCAATCATGTGACTCCGTGG 19515  
 Qy 2271 -----ArgValProLysArgProIleIleAsnThrPro 2281  
 Db 19516 TGAATGTGACCTGACCTGACCCCAATCAAGAGTCCCAACGCGCATCTCAACACACC 19575  
 Qy 2282 ValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysPro 2301  
 Db 19576 GTGGTGAAGCATCAAGCTTCATGATGATGAGAGCTTCTGCCCGGGCCCTGGAACAAACC 19635  
 Qy 2302 ValThrValGlnPheArgLeuLeuGluThrGluGluArgThrLysProIleCysValPhe 2321  
 Db 19636 GTCAAGGCGAGGTCCTCGCTGTGAGACCTGACATGCGCCCTCAAGGCTTCGGGCTGAAG 19695  
 Qy 2322 TrpAsnHisSerIleLeu----- 2327  
 Db 19696 TGGAACTCAATCAATCTGTGAGCTGACATGCGCCCTCGCCCTCAAGGCTTCGGGCTGAAG 19755  
 Qy 2327 ----- 2327  
 Db 19756 GTCCAGAGCCCTGACATGCTCAACCTCTGTCTCCCTGACCCATGCTTCTCAACAGAG 19815

Qy 2328 ValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGlyValValPheArgAsnGluSer 2347  
 Db 19816 GTCAAGTGGACAGAGTGGCTGTGTGGCCAGAGGGTGTGAAGTGTCTTCCGCAATGAGAGC 19875  
 Qy 2348 HisValSerCysGlnCysAsnHisSerThrSerPheAlaValLeuMetAspValSerArg 2367  
 Db 19876 CAGTCAGCTGGCCAGTGGCAACCAATAGAGAGCTTGCTGTCTCATGACGTTTCTCGG 19935  
 Qy 2368 ArgGlu----- 2369  
 Db 19936 CGGAGAGTGGGGCCCAAGGGGAGCTGACAGAGCCGTGGGTGGGACCCAGGGCAGGGGG 19995  
 Qy 2370 -----As 2370  
 Db 19996 CTGGTGTCTCAGGTCTGCGCTTCTTAATTCCTTGAGCCCTTCGCACTTACTTGAGAA 20055  
 Qy 2370 nGlyGluIleLeuProLeuLysThrLeuThrValAlaLeuGlyValThrLeuAlaI 2390  
 Db 20056 TGGGAGATCTCGCCACTGAAGCACTGACATACGTGGCTTGAAGTGTCACTTGGCTGC 20115  
 Qy 2390 AlaLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisG 2410  
 Db 20116 CTTTCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 20175  
 Qy 2410 YlleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyT 2430  
 Db 20176 CATCCACGTAACTGACAGCTGCTCGGGGCTGAGCTGAGTGTCTTCTCTTCTGGGAAT 20235  
 Qy 2430 eAsnGlnAlaAspLeuPro----- 2436  
 Db 20236 CAACAGGCTGACCTCTCTGAAGATGTCCTTACGCGAGAAATGTGCCACTTCTTC 20295  
 Qy 2436 ----- 2436  
 Db 20296 AGGCGCCCTCCCAAGGCCCCCACTGGACACCCCTGCTCTGACCATGAATCTAATAAG 20355  
 Qy 2436 ----- 2436  
 Db 20356 TGCCTAGTCAGACCTGAGCCAGAGGTTCTCTTCTGTGTGCTCCCCGGGATCCCCAG 20415  
 Qy 2436 ----- 2436  
 Db 20416 CACTGCTGTGGCCAGAGCTTCCCTGAAGCATTTCCAAACACCAAGGCCCTCTCCAT 20475  
 Qy 2436 ----- 2436  
 Db 20476 GCCTGAACCGAAGCAGAGCTGTGCTTGGGGGGGCCCCGGTGTGACCTGCGCTGGCC 20535  
 Qy 2437 -----PheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysT 2453  
 Db 20536 TGGGCCCTGAGTTGCTGTCAAGTATGCAATCTGTGTGCACTTCTGTATCTTGTGCA 20595  
 Qy 2453 hPheSerTrpAlaLeuLeuGluAlaLeuHisIleLeuTyrArgAlaLeuThrGluValArg 2473  
 Db 20596 CTTTCTGAGGCTGTGTGTGAGGCTGTGACCTGTGACCGGACCTTCACTGAAGTCCGG 20655  
 Qy 2473 sPValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheI 2493  
 Db 20656 ATTCAAACACCGGCCCAATGCGCTTCTATCAATGCTGGGCTGGGCGTGGCTTCA 20715  
 Qy 2493 lThr----- 2494  
 Db 20716 TCAAGATATCCCAACCAATCCCACTTGGGGTCCCAATCCCTGGGTCCACTTTGT 20775  
 Qy 2494 ----- 2494  
 Db 20776 GCCATGTCTCTCAACCAATACAGAGCCCTGAAGGCCCAATCCCAATGCCCAAGCCG 20835  
 Qy 2494 ----- 2494  
 Db 20836 CTTTATTCACAGGTGTCCCTGTGTAAACCAAGCTTGAAGCCGACCAACCAAGGCGTCA 20895



QY 2494 ----- 2494  
Db 20896 CTCCTTATTCGAGGAGATCGGTAGGGCCGATGTGGGAGAGAACCTTTTCATGCTT 20955  
QY 2494 ----- 2494  
Db 20956 CATGCTGAGCCCTGTGAGCCCACTGCGGACCCCTTACCTCCAGGCCCTCATCAACC 21015  
QY 2495 ----- 2495  
Db 21016 CTCACATGCTCCGCTGTCTCATGCTCCAGGGCTTA-GCCGTGGGCTTGGACCCGAGG 21074  
QY 2504 1YTYRGLYANP-ROAAPHNECYETRPLEUSER11ETRYAPPHRLEU11ETRPSER-PHEA 2524  
Db 21075 GCTACGGGAGACCTGACCTTCTGCTGCTCTCATATGACAGCTCATCTGAGATTG 21134  
QY 2524 1AG1YPROVAL1ALPHEA1VAL1SER----- 2532  
Db 21135 CTGGCCCGGTGGCTTTGGCTTCGCTCCGTAGTGTAGCAGGTGGGTGGGTCTACCTGT 21194  
QY 2532 ----- 2532  
Db 21195 GGCCCTCTTGTGCTGCTCTCTGCTGTGGGCGGGTGTCTCAGGACTCCCTTTAGGA 21254  
QY 2532 ----- 2532  
Db 21255 ACAGCTGAGGCCACGGGGCCCTGTGGGCTGGGTGAGAGCTTTGTCTCCCACTGAGCACC 21314  
QY 2533 ----- 2533  
Db 21315 CCATGCCCTGTGCGCCCTGCTTACATAGTGTCTTCTCTGATACATCTGGGGGCGGCGCT 21374  
QY 2545 ETCYAA1AALAG1NATG1NG1YPHEG1ULYELYSGLY----- 2557  
Db 21375 CCGTGTGCGCCGCGGCGGAGGCTTTGAGAGAAAGTCCGTGTGATAGAGGTGGG 21434  
QY 2557 ----- 2557  
Db 21435 GTGCTGGGCCATGGGACGACACAGCATGGAGGTCTCATGGCCGGACTCATGGCCTGT 21494  
QY 2558 ----- 2558  
Db 21495 CCTATCCCAAGC-TGGGGCTTGCAAGCCCTCTTGCGCTCTCTGCTGTGAGGCA 21553  
QY 2575 hETRPLEU1A1ALEU1SER-VAL1ASN1SERPHR1LEU1PHEH18TYRLEUPHEA 2595  
Db 21554 CGTGGCTGCTGGGCACTGCTCTGTCTCAACAGGACACCCCTTCCACTACTCTTTG 21613  
QY 2595 1ATHRCYABANCYALLE----- 2600  
Db 21614 CTACCTGCAATTCATTCAGGTACTGAGCCACGCTGTGAGAGAGGAGGCACTGGGCT 21673  
QY 2601 ----- 2601  
Db 21674 GTGGATGCTGAATATGACACAGCCGTTGCTGCTCTTGCTGCTGCGAGGGCCCTTATCT 21733  
QY 2606 hELEUSER1YVAL1VA1LEUSER1YGL1VAL1RGLYAL1ALEU1YSEU1A1CYSE1ERA 2626  
Db 21734 TCTCTCTCATATGATGTGTGTAGCAAGAGGTCCGAAAGACACTCAAGCTTTGCTGAGGC 21793  
QY 2626 TGLY1SER1SER-PROA1P-ROA1AL1EUTH1R1LY1SER1R1LEUTH1R1SER----- 2642  
Db 21794 GCAAGCCAGCCCTGACCTGCTGTGACCAACAAATTCACCCCTGAGCTGGTGAAGGAGAC 21853  
QY 2642 ----- 2642  
Db 21854 CAGGGGTCTCAGAGGGCGGTGAAGGAGGAGGAGGACCAAGTGAAGCTTACT 21913  
QY 2642 ----- 2642  
Db 21914 TTTGGCCCACTCCCTTCTCTTTCTCCATCTCTCTGAAAGGTGGAAGGGAGGTG 21973  
QY 2642 ----- 2642

Db 21974 AAATGCTGTGTTGGGCTGGGAGAGATGTTGAAAGGGGTGTGTAGATTGTAATTGA 22033  
QY 2642 ----- 2642  
Db 22034 GGGAGATGGGGCTGAGACAGATTAATCAAGTAATTAAGTGCCTTGACCTGAGAACTCTGA 22093  
QY 2642 ----- 2642  
Db 22094 TGAATTTAGTAATTTTCTGTCTCCAAAGTGTCTGACGAGACAAATTTGTGTCTAT 22153  
QY 2642 ----- 2642  
Db 22154 CTTTACAGAGCATTAATATATAGTAGTCCAGGATCCGTGGGGCATGACGGGTAGACA 22213  
QY 2642 ----- 2642  
Db 22214 GCTGTGGCAGAGTGTGACCACTGCCAGGCTCAATATGGGAGACGGAGAGAGGCGC 22273  
QY 2642 ----- 2642  
Db 22274 CCTCTTTTCAAGCCTGAGCATGTGGAAGTGGAACTTAGGCAAGTTCCCTCAACCTTC 22333  
QY 2642 ----- 2642  
Db 22334 CTTCTTCCAGCCTGGGGCAGCCATCCCACTTACTGACCTCTGTCCCTG 22393  
QY 2643 ----- 2643  
Db 22394 CCTAGTCTTCAACATGCCCCAGCCCTTACGAGAGTGGGGGTGTATACAGGCTTACGGAG 22453  
QY 2661 SP1SER1AG1Y1SER1LEU11IS1SER1R1SER1R1SER1R1Y1SER1R1P1SER1R1P1 2681  
Db 22454 ACTGGGGGGGCTCTGTGACAGGACCAAGTCCGTGGGCAAGGTACGCCACTTACATCC 22513  
QY 2681 1OPHELEU1----- 2684  
Db 22514 CTTCTGTGAGGTAAATCCGAGATGGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGG 22573  
QY 2684 ----- 2684  
Db 22574 CATGTGACCCAGGCGAGCCAGCTGTTGGAGTTGAGAGACACACTGTGGCTGACGT 22633  
QY 2684 ----- 2684  
Db 22634 GGGGGCAAGCTTGATTAGAACTGTAAAGGACCCACAGAGAGAGAGAGATCCAGGGG 22693  
QY 2684 ----- 2684  
Db 22694 AGAGGAGAGACTGGAGCCCTGGGCAAGGGGCGAGGCTGACCCCTCAGCATGTCTCATTC 22753  
QY 2685 ----- 2685  
Db 22754 TTCTTAGGAGAGGTCCGACATGAACCTTGGCGAAGGCGCCCTTGGCTGGGGATTCAG 22813  
QY 2703 1Y1SER1R1PHEU1UG1YGL1N1ASP1G1N1H1ASP----- 2715  
Db 22814 GGAAGCTGTCTTGTGAAGTCAAGACAGAGCATGG-TAGAGACAGAAACGCTTGGCCA 22872  
QY 2715 ----- 2715  
Db 22873 CCAAGAGAGGAGTGTGGCTGCTTTTACTGAAGGTGGGTGAGGTGGGCTGGCTGT 22932  
QY 2716 ----- 2716  
Db 22933 CATCTCTCCCTGGCTCTTAATCTTGAACGAGACTCCGACAGTACCTGTCTTGAAG 22992  
QY 2728 SPASP1N1SER1Y1SER1Y1A1SER1R11AS1SER1R1ASP1SER1R1UG1UG1UG1UG 2748  
Db 22993 ACAGACAGAGTGGCTCTATGCTCTTACCACTATGACAGATGAGAGAGAGAGAGG 23052  
QY 2748 1UG1UG1UG1UG1A1A1PHEPROG1YGL1UG1NG1Y1TPASP1SER1LEU1Y1PROG 2768

Db	23053	AGGAGAAAGAGAGAGCGCGCTTCCCTGGAAGACAGGAGCTGGATTAGCTGCTGGGCGCTGG	23111
Qy	2768	LYIAGIUAAGLUAProLeuHISerThProlys-----	2779
Db	23113	GAGCAGAGAGACTGCGCTCTGACAGTACTCCAAAGGGTGGGCCAGCATGGGGCTGTGGC	23172
Qy	2779	-----	2779
Db	23173	CTTTGGGGCCAGTGGAGGAGACATGTGGCGCTGGGGTTCTTGGAGAGAGACTGGGGTGGT	23232
Qy	2780	-----	2782
Db	23233	GGCCTCTGCGTACTGCGCTCTGTGGGCGCTCATCTTCTTTCCCAACAGAGTGGGG	23292
Qy	2782	LYProGlyProGlyLYValAProTProGlyIAspPheGlyThrThrAlaLYsGLuSers	2802
Db	23293	GCCCAAGGCGCTGGCAAGAGCCCGCTGGCCAGAGACCTTTGGAGACCAAGCAAGAAAGAGAT	23352
Qy	2802	erGIYAengIYAIAProGluGluIUAArgLeuArgLUAasnGlyAAspAlaLeuSersArgLUG	2822
Db	23353	GTGGCAACGGGGCGCCCTGAGAGAGCGGCTGGCGGAGAAATGAGATGCCCTGTCTCGAGAG	23412
Qy	2822	LYSerLeuGlyProLeuProGlyLYSerSerAlaGlnProHISlys-----	2836
Db	23413	GGTCCCTTAGCGCCCTCTTCCAGGCTCTTCTGCGCAAGCTCAAAAGGTAGTGGGCAACC	23472
Qy	2836	-----	2836
Db	23473	CCAGCTGCGCAGCTCCCGCTAGTCAGACAGCTCATTACTCATTTCTGTGGCGGCAAC	23532
Qy	2837	-----	2849
Db	23533	TCACAGCCCGCGCCCGGCCCAACAGGCAATCTTTAAGAAAGTGTCTGGCCACCATCAGAG	23592
Qy	2849	IULYSerSerLeuLeuArgLeuProLeuGluGlnCYthrGlySersSerArgGLYSers	2869
Db	23593	AGAAAGACAGCTCTCTGCGGCTCCCGCTGGAGCAATGCAAGAGTCTTCCCGGGGCTCTCT	23652
Qy	2869	erAlaSerGluGlySersArgGLYGLYProProProArgProProProArgLInSerLeuG	2889
Db	23653	CCGCTAGTGAAGGCGACGCGGGGAGGCCCGCTCCCGCCACGCGCCCGGACAGAGCTCC	23712
Qy	2889	IUGluGluIleuAsnGlyValMetProIleAlaMetSerIleLYsAlaGlyThrValAspG	2909
Db	23713	AGGAGCAGCTGAAGGGGTCTAGTCCATCGCATGACATCAAGGACAGGACGCTGGATG	23772
Qy	2909	IUAepSerSerGlySersGlu 2915	
Db	23773	AGGACTGCTCAGGCTCCGAG 23792	
RESULT 5			
US-09-737-149-1			
; Sequence 1, Application US/09737149			
; Patent No. US20020077466A1			
; GENERAL INFORMATION:			
; APPLICANT: Spaderna, Steven K			
; APPLICANT: Quinn, Kerry B.			
; APPLICANT: Shimketers, Richard A.			
; APPLICANT: Muralidhara, Padiganu			
; APPLICANT: Syntek, Kimberly A.			
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same			
; FILE REFERENCE: 15966-620 CIP			
; CURRENT APPLICATION NUMBER: US/09/737,149			
; CURRENT FILING DATE: 2001-06-15			
; PRIOR APPLICATION NUMBER: 60/170,564			
; PRIOR FILING DATE: 1999-12-14			
; PRIOR APPLICATION NUMBER: 60/173,165			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: 60/173,362			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: 60/173,544			
; PRIOR FILING DATE: 1999-12-29			
; PRIOR APPLICATION NUMBER: 60/174,404			

[illegible]

OY		162	SerProGluGluSerLeuGIyglIyargArgLyAargAsnValAsnThrAlaPro-----	179
Db		664	TTGCCCGGAAGCCCGGGCCGGCGCGCAGACGGCCCGGGCGGCAGAGGG	723
OY		160	-----GlnPheGlnProPserTyrGlnAlaThrValProGluAsnGlnProAlaGly	197
Db		724	AGCTCAAGATTCCCATATGCCAACTCACAGGTGGCGTTTGTAACAAGAACCGCGGACC	783
OY		198	ThrProValAlaSerLeuArgAlaIleAspProAspGluGluAgIyArgrLeuIu	217
Db		784	ACCCTCATCTCCACGTGACAGCGCACTACACATTCAGAGGGGAGAGAGAGCGCTGACG	843
OY		218	TyrrrMetAspAlaLeuPheAspSerAargSerAsnGlnPhePheSerLeuAspProval	237
Db		844	TATTACATGAGAGGGCGCTGTCACACAGAGCGCTCCGGGGCTCATCTCCGATCATGACTCTGC	903
OY		238	ThrgIyAlaValIThrThrAlagluGluLeuAspArgGluThrLysSerThriItyAlaPhe	257
Db		904	ACGGCGCGCGTAGACACGGACAGCGTACTGACCCCGAGAACCAAGAAGACGACCTCTC	963
OY		258	ArgValIThrAlaglInAspHisgelyMeTProAragrSeralaleuAlaThrLeuTrille	277
Db		964	AGGGTGAAGCCGTGAGCTACAGTACGCCCGCGCGCTCGGCACCACTACATCATCTCTC	1023
OY		278	LeuValIThrAspThrAsnAspHisAspProvalPheGluGlnGlnGluITyrrLysgIuser	297
Db		1024	TTGGTCANAAGACCAACAGACACAGCCCGCGTTCGAGAGTCCGAGTATCCGACGGCC	1083
OY		298	LeuArgGluAsnLeuGluValGlYTyrrGluValLeuThrValArgAlaIThrAspGIyAsp	317
Db		1084	GTCGCGGAGAACTGTGAGTGGCTACAGAGGTGTCACATTCGCGCCAGAGACCGGCAC	1143
OY		318	AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProser	337
Db		1144	TCGCCCATCAACGCCCACTTCGCTTACCGCGGTGGGGCGCGCTGC-----	1191
OY		338	GluValPheGluIleAspProArgserGIyValIleArgTrArgGlyProValAspArg	357
Db		1192	GACGCTTCCAGCTCAACAGAGAGCTCTGGCGTGTGAGACACACGGCGCGTCTGACCGG	1251
OY		358	GluGluValGluSerTYrrGlnLeuThrValGluAlaSerAspGlnGlyAragAspProGIy	377
Db		1252	GAGAGGGCGCGCAGTACCACTCTCGTGTGAGGCGCAACGACACAGGGCGCAATCCGGGC	1311
OY		378	ProArgSerThrThrAlaAlaValPheUserValGluAspAsnAspAsnAlaPro	397
Db		1312	CCGCTCAATGCCAGCGCCACCGGTATCACTCAGAGTGAGAGACAGAAAGACATCACTCCC	1371
OY		398	GlnPheSerGluTyAargTyrValValGlnValArgIleAspValIThrProGIyAlaPro	417
Db		1372	CAGTTCCAGCGAGCAAACTACGTGGTCCAGGTGCCGAGGACGCGGGGCTCAACACGGCT	1431
OY		418	ValLeuArgValIThrAlaSerAspAragAspLybelSerAsnAlaValValHISrySer	437
Db		1432	GTGCTGCAAGTGCAGGCGCACCGACCGGGACCAAGCGCGCAATTCATCACTACGC	1491
OY		438	IleMeSerGIyAsnAlaAargGIyGlnPheTyrLeuAspAlaGlnIThrGlyAlaLeuAsp	457
Db		1492	ATCTCTCAGCGGGAACTGGCGCGCGCACTTACTCTGCACTGCTGAGCGGGATCTTGAT	1551
OY		458	ValValSerProLeuAspTyrGluIThrThrLysgIuserIThrLeuArgValArgAlaGln	477
Db		1552	GTGATCAACCCCTTGATTCGAGAGATGCCAGAAATCTCGCTGAGACATTAAAGCCCAAG	1611
OY		478	AspGIyGIyAargProProLeuSerAsnValSerGIyLeuValIThrValGlnValLeuAsp	497
Db		1612	GATGGGGCGCGCCCCCGCTCATCAATCTTCAAGGGGGGAGTGTGTGTGAGATCTGGAT	1671
OY		498	IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer	517
Db		1672	GTCAAACAGCAACGAGCTTATCTTTGTAGAGACCCCTTCAGGCGCACGATCTCTGAGAT	1731

QY	518	ValProLeuEngLYrYrLeuValLeuHisValGlnAlaIleAspAlaGlyAspAsn	537
Db	1732	GGGCCCCCTGGGGCTACCCCTGGGTGACATTCAGCGGGTGGACCGGGACTCTGGAGAGAAC	1791
QY	538	AlaArgLeuGluYrYrArgLeuAla-----GlyValGlyHis---	549
Db	1792	GGCGGGCTGCACATATCGCTGTGTGGACAGGCGCTCCACCTTCTGGGGGGGGGAGCGCCT	1851
QY	550	-----AspPheProPheThrIleAsnAsnGlyYrThrGly	560
Db	1852	GGGCTTAAGAAATCTGGCCCCCAACCCCTGACATTCCTCCCTTCAGAGATCCACAAAGCTCCGGT	1911
QY	561	TrpIleSerValAlaIleGluLeuAspArgGluGluValAspPheYrSerPheGlyVal	580
Db	1912	TGGATCAACAGTGTGTGGCCGAGCTGGACCGCCAGAGAGTGGAGCACTACAGCTTGGGGGTG	1971
QY	581	GluAlaArgAspPheIleGlyYrThrProAlaLeuThrAlaSerAlaSerValThrVal	600
Db	1972	GAGGGGGTGGACCAACGGCTCGCCCCCATAGCTCTCCACCAAGCTGTCCACCGGTG	2031
QY	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluYrThrValArgLeuAsn	620
Db	2032	CTGGACGTGAATGACAAAGACCCGGGTGTCCACGACGCCACTACAGACTTCGCTGAAT	2091
QY	621	GluAspAlaAlaValGlyYrThrSerValValThrValSerAlaValaAspArgAspAlaHis	640
Db	2092	GAGAGTGGCGGCGGTGGGAGACAGCTGTCTGACCTTCGACGGCCCCGACCGTGAAGCCAAC	2151
QY	641	SerValIleThrYrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660
Db	2152	AGTGTGATTCCTACCAAGCTCCACGGGGGCAACCCGGGAACCCCTTGGCATCAGACAGC	2211
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspYrYrYrIleGluArgGln	680
Db	2212	CAGAGAGGGGGGGGCTCATACCTTGGCGCTTCTTGACTACAAAGCAAGACAGACAG	2271
QY	681	TyrValLeuAlaValIleThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValAla	700
Db	2272	TACCTGTGGGGGTGAGACGATCCGAGCGGCAACGGGTGCGACACTGGCGCATGTCTCATC	2331
QY	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyYrThrValAsn	720
Db	2332	AACGTACTGTAGTCCAAACCCCAAGGCGTGTCTTCAGAGCTCCCATATCACAGTAGT	2391
QY	721	ValAsnGluAspArgProAlaGlyYrThrValValLeuIleSerAlaThrAspGluAsp	740
Db	2392	GTCAGTGTAGGACAGGCGCTGTGGGCACTCATTTGGCTACCTCAGTGGCAACGATGAAGAC	2451
QY	741	ThrGlyGluAsnAlaArgIleThrYrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2452	ACAGAGAAAGATGCCGCACTACCTAAGTATTCAGGACCCCGCGCGAGTTCGGCAAT	2511
QY	761	AspAlaAspThrGlyAlaValIleThrGlnAlaGluLeuAspYrGluAspGlnValSer	780
Db	2512	GACCCCGCAAGTGGCACCATGTACACCATATGAGAGCTGACATRTAGAACCAAGTGGCC	2571
QY	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnYrSerAspThrThrYr	800
Db	2572	TACACGCTGACATCATAGCGCCAGAGCAACGGGCACTCCGAGAAATACAGACCAACACC	2631
QY	801	LeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerYr	820
Db	2632	CTAAGAGATCCATCCTCGATGCCAAAGACAAATGACACCCAGATTCGTGGAGATTTCTAC	2691
QY	821	GlnGlySerValYrGluAspValProProPheThrSerValIleGlnIleSerAlaThr	840
Db	2692	CAGGGTTCATCTTTGAGATGTCCACCTTCGACCAAGCATTCCTCCAGGTCTCGACAGC	2751
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheYrYrThrPheGlnGlyValaAspAsnGly	860
Db	2752	GACCGGGAACCTCAGGTCCCAATGGAGCGGTCTGTGACCTTCCAGGGGTGGAGACAGCGC	2811
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuAlaGlyLeuAsp	880

Db 2812 GATGGGACCTTCACTGACGACCCGCTCGGTGATTCGACACCGACGCGCGCTGAC 2871  
 Qy 881 ATGGUAsnValAlaGlnIleValLeuArgAlaTyrAlaValAspIleGlyMetPro---- 899  
 Db 2872 CGGGAAGATGGCCGTGACCAACTTGGGCTCTGGCTGTGATGGAGGAGTCCACT 2931  
 Qy 900 ProAlaArgThrPrometGluValThrValThrValLeuAspValAsnAspAspPro 919  
 Db 2932 CCCCTTAGCGCCTCGGAGAAATCCAGGTGACCATCTTGGAATATATGCAATATGCC 2991  
 Qy 920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 2992 ATGTTTGAAGAGCAACTGAGCTGTTGTGTGAGAGAGAACCAACGAGGGGTGCTG 3051  
 Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3052 GTGGCAAGATTGTGTCTAACGACCTGTATGAAAGCCCTATATCCAGATCATATATCAG 3111  
 Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3112 ATGTGGAGAGGAGCAATGCGGCAATTTCTTCACCTGGAACCTGCTCAACGGGAGCTGCG 3171  
 Qy 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 3172 GGCATGTGAGCTGACCTTGAAGTCCGCGGAGATATGTGTGTGTGTGTCGACGCGCACG 3231  
 Qy 1000 SerAlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAsnAsp 1019  
 Db 3232 TCGGCTCCGCTGTGAGCCGAGCCACGCTGACATCTTCTGTGACCAAGATGACAC 3291  
 Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSer 1039  
 Db 3292 CCGCCCTGTGTGCGCCACTTCAGATCTCTTCAACAATATGTCAACAAGTCAAC 3351  
 Qy 1040 SerPheProGlyGlyValIleGlyArgValProAlaIleAspProAspIleSerAspSer 1059  
 Db 3352 AGTTTCCCAACCGGCTGATCGGCTGATCCGAGCCATGACCCGACGTCGACGAC 3411  
 Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuAsnAlaSerThr 1079  
 Db 3412 CTCACTACACCTTCGTCGACGGGCAAGAGCTCGCTGTGTGTGTGACCCCGGCAAG 3471  
 Qy 1080 GlyGluLeuLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSer 1099  
 Db 3472 GGGGAATGAGCTGACCGCCGACCTGACCAACACCGGCTGAGGGGCTCATGAGAG 3531  
 Qy 1100 ValLeuValSerAspGlyValIleSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db 3532 GTGTCTGTCTGATGGCATCCAGGGTCACGGCTTCTGACCCCTGCGTCACCATC 3591  
 Qy 1120 IleThrAspGluMetLeuThrIleSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 Db 3592 ATACCGAGACAGATGTCGACCAACAGATCACTGCTCGCTGAGAAACATGTCCAGAG 3651  
 Qy 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
 Db 3652 AAGTTCTGTGCGCGCTGCTGCGCTCTCTGTGAGGGGGTGGCCCGCTGCTGTCCACC 3711  
 Qy 1160 ProProAspIleValValIlePheAsnValGlnArgAspThrAspAlaProGlyGlyIle 1179  
 Db 3712 ACCAAGACACGCTTCTGCTTCACAGTCACAGACACCAAGCTC--AGCTCCAAC 3768  
 Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhe 1199  
 Db 3769 ATCTGAACGTGACCTTCTCGGCGCTGCTGCTGCGGCGCTCGCGGCTC-----CA 3822  
 Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 Db 3823 TTCCCGTGGAGACCTGACGAGACAGATCTTACCTGATGAGCGCTGTCGACCACTC 3882  
 Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239

Db 3883 TCCACGACGCGGTGCTCCCTTCGACGACCAATCTGCTGCGGAGCCCTGCGAGAAC 3942  
 Qy 1240 TyrMetArgCysValSerValLeuAspPheAspSerSerAlaProPheIleAlaSerSer 1259  
 Db 3943 TACATGAAGTGTGCTCCGTCTGCGATTTCAGACCTCCGCGCTTCTCTCACTCCACC 4002  
 Qy 1260 SerValLeuPheArgProIleIleAspValGlyGlyLeuArgCysArgCysAspProGly 1279  
 Db 4003 ACCGTGCTTCCGCGCCATCCACCCATCAACCGGCTGCGCTGCGCTGCGCGCGCC 4062  
 Qy 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
 Db 4063 TTCACCGGAGACTCTGAGAGCGAGATCACTCTCTCACTCCGACCCGCGCGCGCC 4122  
 Qy 1300 HisGlyArgCysArgSerArgGlyGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
 Db 4123 AACGCCCTCTCCGAGCCGAGCCGAGGCGGCTACACCTCGAGTCTTGAGAGACTTCACT 4182  
 Qy 1320 GlyGluIleCysGlyValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsn 1339  
 Db 4183 GGAAGACACTGTGAGGTGATGCCCGCTCAGGCGCTGTGCCAACGGGGTGTGCAAGAAC 4242  
 Qy 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheIleCysAspCysProSerGlyAsp 1359  
 Db 4243 GGGGCACTGTGTAACCTGTCAATGCGGCTTCCACTGTGTGTCTCTGCTGCTGAG 4302  
 Qy 1360 PheGluIleProTyrCysGlnValThrIleArgSerPheProAlaIleSerPheIleThr 1379  
 Db 4303 TATGAGAGCCCTTATGTGAGTACACACAGAGCTTCCGCGCCAGTCTTCTGCTCAC 4362  
 Qy 1380 PheArgGlyLeuArgGlnArgPheIlePheThrLeuAlaLeuSerPheIleThrIleGlu 1399  
 Db 4363 TTCCGGGCGCTGAGACACGCGCTTCCACTTCACTCTCCCTCCACCTTGCACATCAGAA 4422  
 Qy 1400 ArgAspGlyLeuLeuLeuLeuThrArgGlyValArgPheAsnGluIleAspPheValAlaLeu 1419  
 Db 4423 AGGAAGCGCTGTCTCTTCAACCGCGCTTCAATGAAAGACAGACTTCACTGCGCTG 4482  
 Qy 1420 GlnValIleGlnGluValGlnLeuThrPheSerAlaGlyGluSerThrThrThrVal 1439  
 Db 4483 GAGATCGGAGACAGACAGAGTGAAGTCACTTCTGTGACGGGAGACAAACAGACCGTG 4542  
 Qy 1440 SerProPheValProGlyGlyValSerAspGlyGlnIleThrIleThrValGlnLeuIleTyr 1459  
 Db 4543 GCACCGAAGTTCACAGTGTGTGATGACGCGGCGGTGCACTGTGTGAGGTGAGTAC 4602  
 Qy 1460 TyrAsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIleVal 1479  
 Db 4603 TACACAAAGCCCAATATTTGGCCACTGGGCTGCTGCCATGGGCTCCGGGGAAGATG 4662  
 Qy 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 Db 4663 GCGGTGTGAGCAGTGAATGTGTGACAAACAGTGGCTGTGTGCGCTTGGAAAGAGATC 4722  
 Qy 1500 GlyAsnTyrSerCysAlaIleGlnGlyThrGlnGlyGlySerIleGlySerLeuAspLeu 1519  
 Db 4723 GGAACCTACAGTGTGCGCCGAGGCACTCAACCGGCTCAAAAGTCCCTGTGATCTG 4782  
 Qy 1520 ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet 1539  
 Db 4783 ACCGCGCTCTACTCTCTGGGGGTGTCCCAACCTGCGCAAGAGATTCCTCACTGACAAAC 4842  
 Qy 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
 Db 4843 CGGCAAGTGTGTGAGTGTGAGGAGAACTGTGACGTGACGCGGAAAGATGTGACATGCGC 4902  
 Qy 1560 AspPheIleAlaAsnArgGlyThrValProGlyCysProAlaIleValAsnValCysAsp 1579  
 Db 4903 GGAATCATGCGCAATAGGACCCGGAAGGCTGCGCTCGAGGAGACTTCTGCGAT 4962  
 Qy 1580 SerAsnThrCysHisAsnGlyGlyTyrCysValIleAsnIleThrPheAspSerCysGlu 1599  
 Db 4963 GGAAGCGGTGTCAAAATGAGGACCTGTGTCAACAGTGAATATGTATCTGTGTAG 5022

QY 1600 CysProLeuGlyPheGlyGlySerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTCCATCTCCGATTCGGCGGAGAACTGTAGACAAAGCCATGCTCAACCCCGACGCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTTPHIsGlyLeuSerLeuProHisSerGlnProTTPYr 1639  
 DB 5083 AGCGGTGAGAGGCTGTCTGTCTGTAGAGTGAACCTGAACATCATCTCTGTGTGCTGTGTAC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 DB 5143 CTGGGGCTCATGTTCGGGACCCGGAAAGAGACAGAGGCTTCTATGTAGAGGCAACAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisIleValMetLeuSerValGlnGly 1679  
 DB 5203 GGGGCCACACAGCTTGTGGCTCCAGATCTGAAACATCACTCAAGTTTGAAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuAlaGlnLeuProGlyValArgAlaAsnAspGlyAspTrp 1699  
 DB 5263 GGCCCTCCGATGTGAGTCCGTGATGTCCGGGTTCGGGTGACCGAGGGGAGTGG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlnHisAlaIle 1716  
 DB 5323 CACCACTCTCTGATCAAGCTGAAAGATGTTAAGAGAGACAGTGAAGTGAACCTGTGTC 5382  
 QY 1717 ---LeuSerPheAspTrpGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
 DB 5383 ACCATGACCTTGAGTACTATGGAGTGAACAGAAACAGAGATATCGGGGCAATGCTTCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
 DB 5443 GGGCTGACGGTGAAGAGCGTGTGTGTGAGAGGCGCTGAAAGACAAAGTCTCCGTGCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsn 1775  
 DB 5503 CGTGAATTCGAGGCTGATGACAGAGAGTGAAGATGGGGGAGCGCCCAACAGTCCGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
 DB 5563 ACCCTGAACATGAAACACCACTCAAGTCAAGGTGAAAGACGCGTGTATGTGAGCAGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyr 1815  
 DB 5623 CCTGTACTCTGACGACCTGTCTCCCAATAGCGCTGCCACACCCCTGGAGAGACTAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 DB 5683 AACTGTGCTGTGTGACAAAGGTACCTTGAAATTAATGTGTGATGCTGTCACTGAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgGlyAspSerAlaProHisGlyTyrThrCys 1855  
 DB 5743 CCTGTGAACATGCGGGCGCTGTGCGCTCCCGGCTCCCGGAGGGTAACTACGTGTC 5802  
 QY 1856 GlnCysProProAsnIleLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
 DB 5803 GAGTGTGGGCGCAAGTCACTACCGGCGGTACTGTGAACAAACTCAACCTTCGTGCCCC 5862  
 QY 1876 ArgGlyTyrTrpGlyHisAspProThrCysGlyProCysAsnCysAspValSerIleGlyPhe 1895  
 DB 5863 AAGAGGTGTGGGGGAGACCCCTGTGTGACCTGTGCACTGTGCTCAAGCAAGGCTTT 5922  
 QY 1896 AspProAspCysAsnIleThrSerGlyGlnCysHisCysValGlyLeuAsnIleTyrTrpPro 1915  
 DB 5923 GATCCCGACGTGAATTAAGACCAACGCGCAAGTGCACATGCAAGAGATTAACAAGCTC 5982  
 QY 1916 ProGlySerProThrCysIleLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 DB 5983 CTAAGCCAGAGACCTGTGTGCTGTGCACTGTGCTCCCAATGAGCTTCCACAGCGCACT 6042  
 QY 1936 CysAspProGlnAspGlyGlnCysProCysValProGlyValIleGlyArgGlnCysAsp 1955  
 DB 6043 TCGACATGCGCAACCGGGCAGTGTGCTCAAGCCCGGCTCATGGCGCGCAAGTGCAC 6102

QY 1956 ArgCysAspAsnProPheAlaGlnValThrThrAsnGlyCysGlnValAsnIleAspSer 1975  
 DB 6103 CCTGTGACCAACCCGTTTGTCCAGGTCACACAGCTCTGGGTGTGAATGATTCATCAATGCC 6162  
 QY 1976 CysProAlaAlaIleGlnAlaGlyIleTTPProAspGlyThrArgPheGlyLeuProAla 1995  
 DB 6163 TGTCCCAAGCATTTTGAAGCCCGGCAATCTGTGGCCACAGACCAAGTTCGGGACGCGGCT 6222  
 QY 1996 AlaAlaProCysProLeuGlySerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 DB 6223 GCGGTGCCATGCTTAAGAGATCCCTTGAAATGCGGTGCACATGCAAGCGGGAGAAAG 6282  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuGlyGly 2035  
 DB 6283 GGTGGCTGTGCCCCAGAGCTTTTAACGTGACCAACATCTCTGTGTGACCTCAAGGGCC 6342  
 QY 2036 PheAlaGlnValGlnGlnArgAsnGlySerGlyLeuAspSerGlyValArgSerGlnGlnLeu 2055  
 DB 6343 ATGAATGAGAACTGAGCGGCATGAGACGCAAGTGTGACGGCGCCAGGGCCCTGAGCTG 6402  
 QY 2056 AlaLeuLeuLeuAlaAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIle 2075  
 DB 6403 GTGAGGGCGCTGTGCGATGTCAACAGACACAGGCAAGCTTGTGGCAATAGCTGCGC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuAlaHisGlnSerThrGlnArgGlyPheGly 2095  
 DB 6463 ACGGCTTACAGCTCTGTGGCGCACAGCTCTTGACAGACGAGGTGCAAGAGGCTTGCAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlyLeuLeuLeuAlaValGlySerAlaLeu 2115  
 DB 6523 CTGGCAGCGACGACAGAGCGGCACTTTCACAGAGACGTCATCACTGGCGAGCCCTTC 6582  
 QY 2116 LeuAspThrAlaAsnIleAspArgHisIleTrpGlyLeuLeuGlnGlnGlnGlyValThrAla 2135  
 DB 6583 CTGGCGCCAGGCCACAGGGCGCGGTGGAGACAGATTCAGCGGAGGAGAGGCGGACGCGCA 6642  
 QY 2136 TrpLeuLeuGlnHisIleTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisIleThr 2155  
 DB 6643 CAGCTGTCCGGCGCTGAGGGCTTAATTCAGCAACAGTGAAGCGCAAGTGTGGCGGAGAC 6702  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
 DB 6703 TACCTGCGGCGCTTGTGATGTGATGTCACGCGCAACATGATCTTGTGTGACATCTTGAAC 6762  
 QY 2176 LysGlyAsnProPheAlaGlyAlaValLeuProArgTyrGlnAlaLeuArgGlyGlyGlnPro 2195  
 DB 6763 AAGTTCACCTTACCGAGAGACAGAGTCCCGGATTCACACATGCATGAAGAGTTCGCC 6822  
 QY 2196 ProAspLeuGlnThrThrValIleLeuProGlnSerValPheArg-----Gln 2211  
 DB 6823 AAGGAGCTGAGTCTCCGTCTCTCCAGCCGCACTTCCAGACCACTGAAGAAAA 6882  
 QY 2212 ThrProProValValArgProAlaGly-----ProGly 2222  
 DB 6883 GAGGCGCCCTGTGAGCGCGGTGCGCGGAGACACCCGCGACACCGCGCGCGG 6942  
 QY 2223 GlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgHisIleProGlyLeuSerGln 2242  
 DB 6943 CCGGCAACAGAGAGAGAGAGCCCGCATGACAGCGGAGGCGACACCTGATGACGCTGCC 7002  
 QY 2243 GlyGlnAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsn 2262  
 DB 7003 CAGTTGCGCGGTGCTGTGTATCATCTTAACCGCACCTGTGGGAGAGTCTGTGCCAGAGCG 7062  
 QY 2263 TyrAspProAspValArgSerLeuAlaValProValAspProIleIleAsnThrProVal 2282  
 DB 7063 TAGACCCCGACCGTGTGAGCTGTGAGTGTGCTCAACGCGCCATATTAATACCCAGAG 7122  
 QY 2283 ValSerIleSerValHisAspAspGlyGlnLeuLeuLeuProArgAlaLeuAspValAspVal 2302  
 DB 7123 GTGAGCAGCGTGTGTATAGAGAGGGGCTCCGCTCCGAGACCCCTGGAAGAGCGCTTC 7182  
 QY 2303 ThrValGlnPheArgLeuLeuGlnThrGlnGlnArgThrIleCysValPheTrp 2322

Db	7183	CTGGGAGAGTTCG	CCCTGCTGGAGGTGGAGGAGCGCAACCAAGCCCTGCTGGGTGTTCTGG	.....	2724
Oy	2323	Aenh1sser1leleuValSer	glYthrTglYTPSer1laArglyCyasgluVala1	.....	2342
Db	7243	AACCACTCCCTGG	CCGTTGGTGGACGGGAGGGTGTGTGCCCCGGGGCTGGACGCTCTGG	.....	7302
Oy	2343	PheArgen1uSer	h1eValSerCyasgluCyasenh1smetThSerPhe1aValleu	.....	2362
Db	7303	TCCAGAACCGGAG	CACATGTGGCTGGCCAGTGCACCCACACAGCCAGCTTTGCGGTGCTC	.....	7362
Oy	2353	MetAspValSer	hArgTgluAsnglylu1leuProleuYrThLeuThTyVal	.....	2382
Db	7363	ATGATATCTCCAG	CGCTGAGAACCGGAGGCTCGCTCGAGATTTGACATATGCC	.....	7422
Oy	2383	AlaLeu	glYVal1ThrLeu1a1aLeuLeuThrPhePheLeuThLeuLeuArg	.....	2402
Db	7423	GCTGTGTCCTTGT	CACTGGACGCCCTGCTGGTGCGCTTGTCTCTTGAAGCTGGTCCGC	.....	7482
Oy	2403	1leleuArgSer	asnglu1h1sgly1leArgArgAsnLeuThra1a1aLeu	glYVala1	2422
Db	7483	ATGCGCGCTCCAA	CTGCACAGCATTCACACAGCACCTCGCGGTGGCTCTTCCCTCT	.....	7542
Oy	2423	GlulnleuValPheLeuLeu	glY1leAsnglu1aAspLeuProPhe1aCyethTyVal1le	.....	2442
Db	7543	CAGCTGGTGTTCG	TATTTGGATTCACACACGAGAAACCCGTTCTGTGACAGTGGTT	.....	7602
Oy	2443	Ala1leleuLeu	h1sPheLeuTyLeuCyThrPheSerTPa1aLeuLeu	glYVala1leu	2462
Db	7603	GCCATTCCTCCCT	CCCACTACATCTACATAGACACCTTGTGCTGGACCTGTGGAGACCTG	.....	7662
Oy	2463	Hi1eLeuTy	raTargalaleuThrgluValArgAspValaAsnThrglyProMetArgPheTy	.....	2482
Db	7663	CATGTCTAACCCCA	CTGCACGAGGTGGCAACATCGACAGCGGGCCATCGGTCTTAC	.....	7722
Oy	2483	Tyrm1eLeu	glYTPglYValPro1aPhe1leThrglyLeu1aVala1leuAspPro	.....	2502
Db	7723	TACGTGCGGGCT	GGGCGATCCCGGCCATTTGTACACAGACTGGCGGTGGCTGGACCC	.....	7782
Oy	2503	GlulglYTy	rglYAsnProAspPheCyStrPleuSer1leTyAspThrLeu1leTPSer	.....	2522
Db	7783	CAGGCGCTACGG	GAACCCCGACCTTCGCTGGCTGCTTCAAGACACCCCTGATTTGAGAC	.....	7842
Oy	2523	Phe1a	glYProVala1aPhe1aValSermetSerValPheLeuTy1leLeu1a1a1a	.....	2542
Db	7843	TTTGGGGGGCCCA	TGCGAGCTGTATATATCAACACAGTCACTTGTCTCTATCTGCA	.....	7902
Oy	2543	Arg1a	AspCyas1a1a1aArgglu1yPheglu1yAslysglyProValSer	glYleu	2562
Db	7903	AAGGTTTCCTGC	CAAGAAAGACACATTAATATGGGAAABAGGATGTCTCTCTGTG	.....	7962
Oy	2563	GlulnProSerPhe	1aValleuLeuLeuLeuSer1aThrTPleuLeu1a1aLeuLeuSer	.....	2582
Db	7963	AGGACCGATTTCC	TTCCTGCTGTCTCATACAGCGCACCTTGCGTGGGGCTGTGCT	.....	8022
Oy	2583	Vala	AsnSerAspThrLeuLeuPheh1sTyThLeuPhe1aThnCyAsnCy1le	glu1y	2602
Db	8023	GTTGAACCGGAT	GCACGTACACTTTCACACTCTTTCGCACCTTCAGCGGCTTACAGGGC	.....	8082
Oy	2603	ProPhe1lePheLeuSer	TyVala1leuSerlysgluValaArglyAsa1aLeuLy---	.....	2621
Db	8083	CCCTTTCGCTCC	CTTTTCCACATCGGTGCTCAACACGAGGTCGCGAGACACTGAAGGGC	.....	8142
Oy	2622	Leu1a	CySerAspTyProSerProAspPro1aLeuThrThrySerThr---leu	.....	2640
Db	8143	GTCCTCGCGGAG	AGAGACCTGCACCTGGAGACTCCGCACACCAAGGGCACCTCTGCTG	.....	8202
Oy	2641	ThreSer	TyTAsnCyAsProSerProTyra1aAsp	glY-----ArgLeuTyrgluPro	2658
Db	8203	ACGGGCTCCCTCA	CTGCAACACCACTTCGATGAGGGCTGCACATGCTCGCACAGAC	.....	8262
Oy	2659	TyrglyAspSer	ra1a1ySer1euh1SerThreAspSer-----	.....	2672

[illegible]





[illegible]

QY	821	GLNGLSERVALYRGIUAAPVALProProPhePheSerValLeuGlnIleSerAlaThr	840
Db	2692	CAGGGTTCACCTTTGAGAGATGCTCACCCCTCGACAGCACTCTCCAGGTCTTGCCACG	2751
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheGlyThrPheGlnGlyValAspAspGly	860
Db	2752	GACCGGGACTCAGGTCCCAANTGGCGGTCTGGTGCACCTTCCAGGGTGGGAGACACGC	2811
QY	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880
Db	2812	GATGGGGACTTCTACATCGACCCACGCTCCGATGTGATTCGACACCGCGCGCTGCAC	2871
QY	881	ArgGluAsnValAlaGlnTyrrValLeuArgAlaTyrrAlaValAspIleGlyMetPro---	899
Db	2872	CGGAGAGATGGCGCGGTGCACACTTTGGGCTCTGGCTGTGANTCGGGGACATCCACT	2931
QY	900	ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro	919
Db	2932	CCCCCTTACGCGCTCCGGGAAATCCAGGTGACATCTTGGCATTAATATGCAATGCCCC	2991
QY	920	ValPheGlnGluAsnAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAla	939
Db	2992	ATGTTTGAAGAAGACGAACTGGAGCTGTGTGTAGAGAAACAACCAAGTGGGTGGGTG	3051
QY	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrrGln	959
Db	3052	GTGGCAAAAGATTCGTGCTAACGACCTCGATGAGAGCCCTAATGCCAGATCATGTATCAG	3111
QY	960	IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr	979
Db	3112	ATTGTGAAAGGGACATGCGGACATTTCTTCCAGCTGACCTCGTCACAGGGGACCTCGT	3171
QY	980	AlaIleuValAspLeuAspTyrrGluAspArgProGluTyrrValLeuValIleGlnAlaThr	999
Db	3172	GCCATGGTGAAGCTGCACCTTGAAGTCCGGCGGAGATATGTGCTGTGTGCACAGCCACG	3231
QY	1000	SerAlaProLeuValSerAlaGlnAlaThrValHisValArgLeuLeuAspArgAsnAspAsn	1019
Db	3232	TGGCTCCGCTGGTGAAGCCGACGACCGATGTCATCTTCTCGTGAACAGAAATGACACAC	3291
QY	1020	ProProValLeuGluAsnPheGlnIleLeuPheAsnAsnTyrrValThrAsnArgSerSer	1039
Db	3292	CCGCTGTGTGCTGCCGACCTTCCAGATCTCTTCAACACTRITGACACCAACAGTCCAC	3351
QY	1040	SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer	1059
Db	3352	AGTTTCCCAACCGGCGATGATCGGTGCATCCCGGCCATGACCCGACGATCAGACGC	3411
QY	1060	LeuThrTyrrSerPheGluArgGlyAsnGlnLeuSerIleuValLeuAsnAlaSerThr	1079
Db	3412	CTCAACTACACTTTCGTCAGGGGACAGAGCTGGCTGTGCTGTGCAGACCCGCCACG	3471
QY	1080	GlyGluLeuLeuLeuSerAlaAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSer	1099
Db	3472	GGGACACTGGACGCTCAGCCGCGACCTGGACAACAACGGCGCGCTGAGGGGCTATGGAG	3531
QY	1100	ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle	1119
Db	3532	GTGTCTGTGTCTATGGCATCCACAGGTGCACGGCTTTCGACACCTTCGCGTGCATCATC	3591
QY	1120	IleThrAspGlnMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu	1139
Db	3592	ATCACGACACAGATGCTGACCAACAGCATCACTCTCGCTGGAGAAACATGTCCAGAG	3651
QY	1140	ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr	1159
Db	3652	AAAGTTCTGTGCCCCGCTGCTGCGCTTGTGTGAGGGGGGTGGCGCGCTGTCTCAC	3711
QY	1160	ProProAspHisValValAlaPheAsnValGlnArgAspHisAspAlaProGlyGlyHis	1179
Db	3712	ACCAAGACACAGCTTCTGTCTTCAAGCTCAGACACACGACGATC---AGCTCCAC	3768
QY	1180	IleLeuAsnValSerLeuSerValGlyIleProProGlyProGlyGlyProProPhe	1199

Db	3769	ATCCGAAAGTACCTTCTCCGGCCGTCGCTGCTGGCCGGCGTCCGGCC-----CAATTTC	3822
Oy	1200	LeuProSerGIuAspLeuGIuArgLeuYrLeuAsnArgSerLeuLeuThrAlaIle	1219
Db	3823	TTCCCGTGGAGAGACCTGCAGAGAGAGATCTACCTGAATCGGACCGCTCTGACCAACATC	3882
Oy	1220	SerAlaGlnArgValLeuProPheAspAspAsnIleCyLeuArgGluProCyGluIuIen	1239
Db	3883	TCCAGCAGACGGCGTCTGCTCCCTTCACACAAACATCTGCTCGGGAGCCCTGCAGAGAC	3942
Oy	1240	TyrMetArgCyValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer	1259
Db	3943	TACATGAAGTGGGTGCTCCGTTCTGCGATTCCGACACTCCGGCCCTTCTCAGACTCCAC	4002
Oy	1260	SerValIleuPheArgProIleHisProValGlyLeuArgCyValCyProProGly	1279
Db	4003	ACCGGCTCTTCCGGCCCAATCCACCCCAACACGGCTGGCGCTGGCCGGCCGGCC	4062
Oy	1280	PheThrGIuAspYrCySerGIuThrGIuValAspLeuCyPyrSerArgProCyGlyPro	1299
Db	4063	TTCAACGGCGAGACTGTCGAGAGCGAGATCAACTCTCTACTCCGACCCCTGGCGCC	4122
Oy	1300	HisGlyArgCyValArgSerArgGluGlyGlyYrThrCyLeuCyAlaAspGlyYrThr	1319
Db	4123	AACGGCCCTCTCCGACGGCCGAGGGCGGCTTACACTCGAGTGGCTTGAGAGACTTCACT	4182
Oy	1320	GlyGIuHisCySerGIuValSerAlaArgSerGIuArgCySerProGlyValCyAlaValAsn	1339
Db	4183	GGAAGACCTGTAGAGTGAATGCCCGCTCAAGCCGCTGTGCCAACGGGGGTGTGCAGAGAC	4242
Oy	1340	GlyGlyYrThrCyValAsnLeuLeuValGlyGlyPheIysCyAspCyProSerGIuAsp	1359
Db	4243	GGGGGCACTGCGTGAACCTGCTCATCGGGCGCTTCACTCGGTGTCTCTCGGGGAG	4302
Oy	1360	PheGIuIysProPyrCySerGIuValThrThrArgSerPheProAlaHisSerPheIleThr	1379
Db	4303	TATGAGAGGCCCTACTGTAGAGTGCACACAGAGACTTCCGCGCCCACTCTTGATCAC	4362
Oy	1380	PheArgGIuLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrTyrGluIu	1399
Db	4363	TTCCGGGGCCGTGAGACAGCGCTTCACTTCAACATCTCCCTCAGCTTGCCACTCAGAA	4422
Oy	1400	ArgAspGIuLeuLeuLeuYrAsnGIuValArgPheAsnGluYrHisAspPheValAlaLeu	1419
Db	4423	AGGAACGGCTTGTCTCTTACACACGGCCGTTCAATGAGAGACAGACTTCATCGCCCTG	4482
Oy	1420	GluValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrThrVal	1439
Db	4483	GAGATCGTGAAGACAGAGGTGAGCTCACTTCTGTGAGCGAGAGACAAACAACCTGTG	4542
Oy	1440	SerProPheValProGlyGlyValSerAspGIuGlnTrpHisIleThrValGlnLeuLeuYr	1459
Db	4543	GACCGAAGGTTCACAGTGTGTGATGACCGGGCGGTGGCACTCTGTGCAAGTGCAGATC	4602
Oy	1460	TyrAsnIysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIleVal	1479
Db	4603	TACACAAAGCCCAATATTGTGGCACTCGGGCTGCCCCCAATGGCGCTCGGGGAAAAATG	4662
Oy	1480	AlaValValThrValAspGIuCySerAspThrGIuValAlaLeuArgPheGlySerValLeu	1499
Db	4663	GCGGTGTGACAGTGTGATGATTGTGACAAACAGTGTGTGCGCTTTGGAAAAAGATC	4722
Oy	1500	GlyAsnYrSerCyValAlaGlnGlyYrThrGlnGlySerIlyIysSerLeuAspLeu	1519
Db	4723	GGGAACCTACAGCTGGCTGGCCCAAGGGCACTCAGACCGGCTCCAAGATCTCCTGATCTG	4782
Oy	1520	ThrGlyProLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet	1539
Db	4783	ACCGGCTTCTACTCTGGGGGGTGTCCCAACCTGCAGAGAGACTTCCAGTGCACAC	4842
Oy	1540	ArgGlnPheValGlyCyMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla	1559

Db	4843	CGGCAAGTTGCTGGGGCTGCATGCGGAACTGTGCATGCGGCAAAATGTGCATGCGC	4902
Qy	1560	AepheiiEaliAaemAemgilyThValProgilyCySProAlalybYbaenValCybaSP	1579
Db	4903	GGATTTCATGCGCAACAAATGGGCAACCGGGAAAGGCTGCGCTGTCCGGAGAACTTTCGGCAT	4962
Qy	1580	SerAenThrCybHiaemgilygilyThCybValaemgintpraphAlaPhaSerCybeIlu	1599
Db	4963	GGAGAGCGGTGTCAAGATGAGAGGCACTGTGTCAACAGGTGAATATGTACTGTGTAG	5022
Qy	1600	CySProLeugilyPhegilygilyYLySerCyValaIngilwEcalAaSPProGlnhiAphE	1619
Db	5023	TGTCACATCCGATTCGGGGGAAAGAACTGTAGGAGGCAAGCTCCACCCCACTCTTC	5082
Qy	1620	LeugilySerSerLeuValAlaTPrhiAgiLyLeuSerLeuProllEserGlnProTPrTy	1639
Db	5083	AGCGGTGAGACCGCTGTCTCTCGAGTGAACCTGAACATCACTCATCTCTGTGCTGCTGAC	5142
Qy	1640	LeuSerLeuMetPheArgThArgGlnhAlaSPgilyValleuLeuGlnAlaIleThArg	1659
Db	5143	CTGGGGCTCATGTTCCGGAACCCGGAAGAGAGACAGCTTCTGTATGAGAGGCCACAGTGT	5202
Qy	1660	GlyArgSerThriIleThLeuGlnLeuArgGlyIyhiAValMetLeuSerValGlyIy	1679
Db	5203	GGGCGCCACCACTTTCGCTCCAGATCTCGAATCCGAACCACTACCTCGATTGAGGTGCCAC	5262
Qy	1680	ThrgilyLeuGlnAlaSerSerLeuArgLeuIuProGilyArgAlaAaenAergilyAaSPTr	1699
Db	5263	GGCCCTCCAGATGTGAGTCCGTATGTCTGCGGGTTCGCGGTGACCGAGGAGAGTGG	5322
Qy	1700	HiAhiAalagIleuAlaLeu-----GlyAlaSerGilygilyProGilyhiAAlle	1716
Db	5323	CACCACTGCTGATCGAGCTGAAGAATGTTAAGAGAGACAGTGAAGAGACATGTGTCT	5382
Qy	1717	---LeuSerPheAaPPrTyGilyGlnGlnArgAlaGlyIyAenLeuGilyProArgLeuHiB	1735
Db	5383	ACCAATGACCTTGGACTATGGAGTGAACAAGAACAGACAGATATCGGGGCACTTCC	5442
Qy	1736	GlyLeuHiAaLeuSerAaenIleThValGlyIyGilyLeuProGilyProAlaGlyIyAla	1755
Db	5443	GGGCTGACCGTAAGAGACCGTGTGTGTGAGAGCCCTCTGAAGCAAGTCTCCGTGCG	5502
Qy	1756	ArgGlyPheArgGilyCybLeuGlnGilyValArgValSerAaPPrThrProGilyValAaen	1775
Db	5503	CGTGATTTCCAGGCTGCATGCAAGAGAGTGAAGTGGGGGAGACCCCAACCAAGTGTGCC	5562
Qy	1776	SerLeuAaPProSerAhiAgiLyIuSerIleAaenValGlnGlnGilyCySerLeuProAaP	1795
Db	5563	ACCTGAACATGAACAAAGCACTCAAGGTCAAGGTGAAGACGGCTGTGATGTGAAGAC	5622
Qy	1796	ProCybaSPSerAaPProCybProAlaAaenSerTyCySerAaenAaPPrPaSPSerTy	1815
Db	5623	CCGTGACCTGAGACCCCTGTGCCCCCAATATGCGCGTCCACAGACGCTGGAGAGACTAC	5682
Qy	1816	SerCybaSPSerCybaAPProGilyTyTyTyGilyAaSPaenCybThAaenValCybaPLeuAaen	1835
Db	5683	AGCTGCTGTGTGCAAAAGGTTACTTGAATAACTGTGTGAATGCTGTCACTGAAC	5742
Qy	1836	ProCybGlnhiAgiIuSerValCybThArgLybProSerAlaProhiAgiTyTyThCyb	1855
Db	5743	CCCTGCGAAGAACAGGGGGCTGTGTGCTCCCCCGGCTCCCGGACGGCTACGTGTGC	5802
Qy	1856	GlyCybAPProAaenTyTLeuGlyProTyTyCybGlyThArgIleAaPPrProCybPro	1875
Db	5803	GAGGTGTGGCCCACTCACTACGGGCTGTGTGTGAACAACCTGCACTTCCGTGCCCC	5862
Qy	1876	ArgGlyTyTyTyTyTyhiAaPPrThCybGilyProCybaAaenCybaAPValSerTyGilyPhe	1895
Db	5863	AGAGGCTGTGTGGGGAACCCCGTGTGTGAACCTGCACTGTGTGCGGTCAAGCAAGCTTT	5922
Qy	1896	AaPPrAaPProCybaenTybThSerGilyGlyCybhiAeCyalybGlnAaenIbTyTyArgPro	1915
Db	5923	GATCCCACTGTATATAGCAACAAGGCAAGTGCACAAATGCAAGAGATTAATCAACATCTC	5982

QY 1916 ProGlySerProThrCysIleuLeuCysAspCysTyrProThrGlyIleuSerArgVal 1935  
DB 5983 CTAGCCAGGAGCACTGTCTGCGCCCTGCGAGCTGCTCCCATGGCTCCCAACAGCCGCACT 6042  
QY 1936 CysAspProGluAspGlyGlnCysProCysValProGlyValIleGlyArgGlnCysAsp 1955  
DB 6043 TGGGACATGGCCACCGGAGGTGTGCTGCGAACCGCGCGTCACTCGCGCCAGTGCAC 6102  
QY 1956 ArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSer 1975  
DB 6103 CGCTGGAGCAACCGCTTGTGCGAGTCCACCGCTGCGTGTATGATCTTCAATGGC 6162  
QY 1976 CysProArgAlaIleGluValGlyIleTyrTyrProArgThrArgPheGlyLeuProAla 1995  
DB 6163 TGTCCCAAGCATTTAGAGCGCGCATCTGTGCGCACAGACAAATTCGGGCGAGCGGCT 6222  
QY 1996 AlaAlaProCysAspProIysGlySerPheGlyThrAlaValArgIleCysAspGluIleAsp 2015  
DB 6223 GCGGTGCGACGCTTAAAGGATCGTTGMAATGCGGTCCGACATGCGAGCGGGAGAG 6282  
QY 2016 GlyTyrLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGluLeuGly 2035  
DB 6283 GGTGTGCTGCGCCCAAGAGCTTTAACTGTACACATCTCTTCTGTGAGACTCTAGGCGC 6342  
QY 2036 PheAlaGluArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnLeu 2055  
DB 6343 ATGAATGAGAAAGCTGAGCGCGCAATGAGACGCAAGTGCAGCGCGCGCTGCGAGCTG 6402  
QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnIleSerAlaGlyTyrPheGlySerAspValLys 2075  
DB 6403 GTGAGGCGCTGCGAGTGTCTACACAGACACGCGGCACTCTTGTGCAATGACGTGCC 6462  
QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaIleGlySerThrGlnArgGlyPheGly 2095  
DB 6463 ACGGCTACAGCTGTGCGGCGCACGTCCTTCAACAGAGAGCTGCGAGAGGCTTGCAGC 6522  
QY 2096 LeuSerAlaThrGlnAspValIleSerThrGlnAsnLeuArgValGlySerAlaLeu 2115  
DB 6523 CTGGCAGCCACGAGACCGCGCATTTCAAGAGAGTCACTCACTGCGAGCGCGCTC 6582  
QY 2116 LeuAspThrAlaAsnIleArgIleSerGlyLeuIleGlnGlnThrGlyGlyThrAla 2135  
DB 6583 CTGGCCCCAGCCACAGGCGCGCTGAGACATTCAGCGAGCGAGCGGCGCACGGA 6642  
QY 2136 TrpLeuLeuGlnIleSerGlyAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgIleThr 2155  
DB 6643 CAGCTGCTCGGCGCTCGAGGGCTACTTCAGCACTGCGACGAACTGCGGCGAGCG 6702  
QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
DB 6703 TACCTCGCGCTTGTCTATCTGTCACCGCCACATATCTTGTCTGTCAATCTTTGAGC 6762  
QY 2176 LysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGlnPro 2195  
DB 6763 AAGTCACTTACGGAGCGGAGGTCGCGCATTCGACACATCCATGAAAGATTCCCC 6822  
QY 2196 ProAspLeuGlnThrThrValIleLeuProGluSerValPheArg-----Glu 2211  
DB 6823 AGGAGACGTGAGTCTCGCTCTTCCAGCGCATCTTCAAGACCACTGAAAGAAAA 6882  
QY 2212 ThrProProValValArgProAlaGly-----ProGly 2222  
DB 6883 GAAAGCCCCCTGTGTAGGCGCGCTGCGCGAGGACCAACCCCGAGACCAACGCGCGGGG 6942  
QY 2223 GluAlaGlnGluProGluGluLeuAlaArgArgGlnArgArgIleProGluSerGln 2242  
DB 6943 CCGGAGACGAGAGGAGGCGCGCGCATCAGAGCGGAGGAGACCTGTATAGCGTGGC 7002  
QY 2243 GlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProIleAsn 2262  
DB 7003 CAGTTGCGCGTCTGTCTCATATTTACCGCACCTGCGGCGAGCTCTGCGCGAGCGC 7062

QY 2263 TyrAspProAspLysArgSerLeuArgValProIlyArgProIleIleAsnThrProVal 2282  
DB 7063 TACGACCCCGAACCGTGCAGAGCTCCGGTTGCTTCAACCGGCCCATATTAATACCCGATG 7122  
QY 2283 ValSerIleSerValIleAspAspGluGluLeuLeuProArgAlaLeuAspLysProVal 2302  
DB 7123 GTAGACACGCTGTATACGACGAGGGGCTCCGCTCCGAGACCCCTGAGAGGCGCGTTC 7182  
QY 2303 ThrValGlnPheArgLeuLeuGlnThrGluArgThrLysProIleCysValPheTyr 2322  
DB 7183 CCGTGGAGTTGCGCTGCTGAGGTCGAGAGCGACCAACAGCTGTCTGCGCTTCTG 7242  
QY 2323 AsnHisSerIleLeuValSerGlyThrGlyTyrPheSerAlaArgGlyCysGluValVal 2342  
DB 7243 AACCACTCCCTGCGCGTGTGTGAGAGGAGGTGTGTGCGCGGGCGCGAGCTCTCG 7302  
QY 2343 PheArgAsnGlySerHisValSerGlyGlnCysAsnHisMetThrSerPheAlaValLeu 2362  
DB 7303 TCCAGAACCGGACACATGTGCTGCGAGTGCAGCCACACAGCAAGCTTTCGCGTGTCTC 7362  
QY 2363 MetAspValSerArgArgGluAsnGlyIleLeuProLeuLysThrLeuThrTyrVal 2382  
DB 7363 ATGATATCTCCAGGCGGTGAGAGAGGAGGTCTGCTCTGATAGATTGTCACTATGCC 7422  
QY 2383 AlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeuLeuArg 2402  
DB 7423 GCTGTGCTCTGTCACTGAGCGCTGCTGTGTGCTGTCTGCTGAGCTGTGTCCGC 7482  
QY 2403 IleuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAla 2422  
DB 7483 ATCTGTGCTCCACCTGCACATTCACAGATTCACAGACCTGCGCGGTGCGTCTTCTCT 7542  
QY 2423 GlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIle 2442  
DB 7543 CAGCTGAGTGTGTGATTTGGATTCACACAGAACCGGCTTCTGTGACACATGGT 7602  
QY 2443 AlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGluAlaLeu 2462  
DB 7603 GCCATCTCTCTCACTATCATCTACATGACACCTTGTGCTGACCTCGTGAAGAGCTG 7662  
QY 2463 HisLeuTyrArgAlaAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTyr 2482  
DB 7663 CATGTCTACCGCATCTACCGAGGTGCGCAACATTCAGACGCGGCGCATGTGCTTCTAC 7722  
QY 2483 TyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspPro 2502  
DB 7723 TACGTGAGGCTGTGGGATCCCGGCATTTGTCAAGAGCTGCGGTGCGCTGAGACCC 7782  
QY 2503 GlnGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrTrpAspThrLeuIleTrpSer 2522  
DB 7783 CAGGCTACGGGAAACCGGACTTGTGCTGCTGTGCTGCTCAAGACCACTGTGATTTGAGC 7842  
QY 2523 PheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAla 2542  
DB 7843 TTTCGGGGCCCATATGAGGCTTTATATCAACACAGTCACTTGTCTATCTGCA 7902  
QY 2543 ArgAlaSerCysAlaAlaGlnArgGlnIlyPheGlyLysGlyProValSerGlyLeu 2562  
DB 7903 AAGGTTCTCTGCAAAAGAAACACCAATATATGAGAAAGGAGATGTCTCCCTGCTG 7962  
QY 2563 GlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuAlaLeuLeuSer 2582  
DB 7963 AGGACCGCATTTCTCTGCTGTCTACAGCGCACCTGCTGCTGCGGTGCTGCTGCT 8022  
QY 2583 ValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGly 2602  
DB 8023 GTGAACCGGATGACATGAGCTTTCATACCTTCTGCGCATCTTCAAGGGCTTACAGGGC 8082  
QY 2603 ProPheIlePheLeuSerTyrValValLeuSerLysGluValArgValAlaLeuLys--- 2621  
DB 8083 CCTTGTCTCTCTTCTTCTCACTGCTGTCTCAACGAGGAGTCCGAGAACCACTGAAAGGC 8142  
QY 2622 LeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThr---Leu 2640

[illegible]

```

! Sequence 749, Application US/10295027
! Publication No. US20030232350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevizi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Ros Biotechnology, Inc.
TITLE OF INVENTION: Methods of diagnosis of cancer, compositions and
TITLE OR INVENTION: Methods of screening for modulators of cancer
FILE REFERENCE: 018501-012500US
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 749
LENGTH: 11389
TYPE: DNA
ORGANISM: Homo sapiens
US-10-295-027-749

Alignment Scores:
Pred. No.: 0 Length: 11389
Score: 8781.50 Matches: 1719
Percent Similarity: 69.42% Conservative: 422
Best Local Similarity: 55.74% Mismatches: 689
Query Match: 56.49% Indels: 254
Gaps: 42
US-09-916-849A-3 (1-2923) x US-10-295-027-749 (1-11389)

QY 13 ProProProProlenleuleuleuleuleuleuleu-ProProProlenleuglyAs 32
Db 10 CCGCCCGCCGCCCTGTGGTCTGTCCTCGGCCGCCGCCG----- 58
QY 32 pginValgIyProCyArSserIeuGLY---SerArgdIyARglISer 47
Db 59 -----ccctgcggcCAnrgggcgTgcgaGacGcgcccgcttgagAcCGCGGTACC 106
QY 48 -----SergIyAlACyalaPrOWetglYTP-----Le 57
Db 109 GGCGAGACCGCGCCCTTGCGCCCTCGGC CGCGCTGTACTACGGGTGGCGCGCTTGC 168
QY 57 uCySProsrSerAlasertAmleutPlleutyRThSer----- 70
Db 169 AGCCCCGGGGCCCCGGGAAGtGCTGTGAACGTGGCGCCGATGAGCGAGAGCT 228

```





Db 2332 AACGTCACTGATGCGCAACACCCAGGCGCTGTTTCAAGCTCCATTACACGTGAGT 2391  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValAlaLeuIleSerAlaThrAspGluAsp 740  
Db 2392 GTCAAGTGAAGACAGGCGCTGTGGGCACTTCACTTGAATCCAGTGCACAGATGAGGAC 2451  
Qy 741 ThrGlyGluAspAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACAGGAGAAATATCCCGCATCACTTACGTGATTCAGAGACCCCGTGGCGCATCCCAAT 2511  
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspThrGluAspGlnValSer 780  
Db 2512 GACCCCGACAGTGGGACACATGACACATGATGAGTGGACATGATGAGACAGGTGGCC 2571  
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrThr 800  
Db 2572 TACACCTGATCCATCAATGCGCCGAGCAACAGGCAATCCCGCAAAATCAGACACACACC 2631  
Qy 801 LeuGluIleLeuValAsnAspValAlaAspAsnAlaProGlnPheLeuArgAspSerThr 820  
Db 2632 CTAGAGATCTCATCTCCATGCAATGCAATGCAATGCAACCCCATGTTCTGTGGATTTCTAC 2691  
Qy 821 GlnGlySerValThrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGCTTCATCTTGAAGATGCTCACTCGACACAGATCTCTCAAGGTCTGTGCCACG 2751  
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheThrThrPheGlnGlyValAspAspGly 860  
Db 2752 GACCGGCACTCAAGGTCCCATATGGCGCTGTGCTGTACACCTTCAAGGTGGGGAGCGAGC 2811  
Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValAlaArgThrLeuArgIleAsp 880  
Db 2812 GATGGGAGACTTCTTACATGAGCCAGCCAGTCCGGTGTATTCGCAACCGAGCGCGCTGGAC 2871  
Qy 881 ArgGluAsnValAlaGlnThrValIleuArgAlaTyrAlaValAspLysGlyMetPro--- 899  
Db 2872 CCGGAGAAATGTGGCCCTGATCAACCTTGGGCTCTGGCTGTGATCGGGAGCTCCCACT 2931  
Qy 900 ProAlaArgThrProMetGluValThrValThrValIleuAspValAsnAspAsnProPro 919  
Db 2932 CCCCTTACGCGCTCGGTGAAATCCAGGTGACCATCTTGGAATATGACATGATGCCCC 2991  
Qy 920 ValPheGlnGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTTTGAAAGAGACCAATGAGCTGTGTTGAGAGAGAACAAACCAATGGGCTCGGTG 3051  
Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGCAAAATTCGTGCTAACACCTGATGAAAGCCCTAATGCCAGATCATGATTCAG 3111  
Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
Db 3112 ATTTGGAGAGGGAGCAATGCGGCAATTTCTTCAAGCTGACCTGTCAACGGGGAGCTGGCT 3171  
Qy 980 AlaLeuValAspLeuAspThrGluAspArgProGluThrValIleuValIleGlnAlaThr 999  
Db 3172 GCCATGTGTAGCTGACTTGTAGGTCCGGCGGAGATATGTGTGTGTGTGAGGCGACG 3231  
Qy 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGCTGTGAGCGGACGACGATGACATCTTCTGTGGACAGATGACACAC 3291  
Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3292 CCGCGCTGTGCGCCATCTTCAGATCTCTTCAACATATGTCCCAACAGTCCACAC 3351  
Qy 1040 SerPheProGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerIleAspSer 1059  
Db 3352 AGTTTCCCAACCGGCGGTGATCGGTGATCCGCGCCATGACCCCACTGTCCAGACAC 3411  
Qy 1060 LeuThrThrSerPheGluArgGlyAsnGluLeuSerLeuValIleuLeuAsnAlaSerThr 1079  
Db 3412 CTCACACTACACCTTGTGAGAGGCGACAGAGCTGCGCTGTGTGTGTGACCCCGCGACG 3471

Qy 1080 GlyGluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluValAlaIleMetSer 1099  
Db 3472 GCGAACTGCACTCAGCGCGGACCTGGAGCAACACCGCGCTGGAGCGCTGATGAG 3531  
Qy 1100 ValLeuValSerArgGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 3532 GTGTGTGTGTGATGAGGATTCACAGCGCTCAGCGCTTCTGTGACCTGTGTGTACACAC 3591  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgGluAspMetSerProGlu 1139  
Db 3592 ATCAAGAGACATGCTGATCAACAGATCACTGATCGCTGAGAAACATGTCACGAG 3651  
Qy 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 3652 AAGTTCTGTGCTCCCGCTGTGGCCCTTCTGTGAGAGGGGGTGGCGCGCTGTGTACAC 3711  
Qy 1160 ProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
Db 3712 ACCAAGACAGAGCTTGTGTCTTCAACGTCCAGAACGACACCGAGCTC---AGCTCCAAC 3768  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyValGlyProProPhe 1199  
Db 3768 ATCTGAACTGATCACTTCTCGCGCTGCTGCTGGCGGCTGCCGCGC-----CAGTTTC 3822  
Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuThrLeuAsnArgSerLeuLeuThrAlaIle 1219  
Db 3823 TTCCCGTGGAGAGACCTGACAGAGAGAGATCTACCTGAATCCGACCTGTGTACACATC 3882  
Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Db 3883 TCCACGACGCGGTGCTCCCTTCCACACACATCTCCTGCGGAGACCCCTGCGCAAGAC 3942  
Qy 1240 TyrMetArgCysValSerValLeuAspPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 3943 TACATGAAGTGTGTCCGTTCTGTGATTCAGACACTCGCGCTTCTCTCAAGCTCAC 4002  
Qy 1260 SerValLeuAspAspArgProIleHisProValGlyGlyLeuAsnArgCysArgProProGly 1279  
Db 4003 ACCGTGCTCTTCCGCGCATCCACCCATCAACCGGCTGCGCTGCGCTGCCCGCGCGG 4062  
Qy 1280 PheThrArgLysAspThrCysGlnThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 4063 TTCACCGCGACTACTGTGAGACGAGATGACCTTCTCTACTCCAGCCCGTGGCGCGC 4122  
Qy 1300 HisGlyArgCysArgSerArgGlyGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 4123 AACGCGCTCGCGGACGCGGACGAGGCGGCTACCTGCACTGCACTGCGAGTCTTGAGAGACTTCACT 4182  
Qy 1320 GlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysLysAsn 1339  
Db 4183 GGAAGACACTGTAGAGTGAATGCCGCTCAGGCGCTGTGCAACGGGGGTGTGCAAGAAC 4242  
Qy 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAsp 1359  
Db 4243 GGGGCACTCGGTGAACTGTCTCATCGCGGCTTCCATCGTGTGTCTCTCCGCGCAG 4302  
Qy 1360 PheGluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Db 4303 TATGAGAGCCCTTACTGTGAGTGAACACAGAGACTTCCGCGCCCACTCTTGTGTAC 4362  
Qy 1380 PheArgGlyLeuAspArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGlu 1399  
Db 4363 TTCCCGGCGCTGAGACAGCGCTTCACTTCACTTCCCTCCCTCAAGTGTGCAAGGAA 4422  
Qy 1400 ArgAspGlyLeuLeuLeuThrThrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeu 1419  
Db 4423 AGAAAGCTGTGCTTCTTACAGCGCGCTTCAATGAGAGACAGACTTCAATGCGCGCTG 4482  
Qy 1420 GluValIleGlnGluGlnValAlaGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
Db 4483 GAGATCGTGAAGAGAGAGTGTCAAGTCACTTCTGTGAGGAGAGACAAACAGACCGTGT 4542

QY 1440 SerProPheValProGlyGlyValSerAspGlyGlnThrHisThrValGlnLeuValTyr 1455  
 Db 4543 GCACCCAGAGGTTCCAGTGGTGTGAGTGCAGCGCGGTGCCTGTGCAGGAGGAGTAC 4602  
 QY 1460 TyrAsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlyVal 1479  
 Db 4603 TACACAGAGCCCAATTTGGCCACCTGGGAGCTGCCCAATGGAGCGGCGGAAAGAG 4662  
 QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 Db 4663 GCGGTGGTGCAGTGGATGATGTGTGACACAAACATGGCTGTGGCTTTGGAAAGAGCATC 4722  
 QY 1500 GlyAsnTyrSerGlyAlaAlaGlnGlyThrGlnGlyGlySerValValSerLeuAspLeu 1519  
 Db 4723 GGGAACTACAGCTGCCTGGCCAGGCGCTCAGACCGGCTCCAAAGAGTCCCTGGAGTCCG 4782  
 QY 1520 ThrGlyProLeuLeuLeuGlyValProAspLeuProGlySerPheProValAspMet 1539  
 Db 4783 ACCGGCTCTACTCTGGGGGGGTGTCCCACTTCGAGAGACTTCCCACTGSCAAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisLeuPheMetAla 1559  
 Db 4843 CGCAGATGGTGGGCTGCAGTGCAGAACCTGTCACTGCAGCAACGCAAAATGTGCATGGCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaValValAsnValCysAsp 1579  
 Db 4903 GGAATATCGCCCAACAATGTGCACCCGGAAGCTGCTGCTGGAGAACTTCTCGCAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGlu 1599  
 Db 4963 GGGAGCGGTGTCAAGATGAGAGCACTGTGTCAACAGGTGGAATGTGATCTGTGTAG 5022  
 QY 1600 CysProLeuGlyPheGlyGlySerCysAlaGlnIleMetAlaAsnProGlnHisPhe 1619  
 Db 5023 TGTCCACTCGGATTCGGCGGGAAGAACTGTGACCAAGCATGTGCTACCCCGAGCTTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProTyr 1639  
 Db 5083 AGCGGGAGAGCGCTGTCTCGTAGTGACTTGAATCATCTCTGTGCTGTGCTGATAC 5142  
 QY 1640 LeuSerIleMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTGGGGCTCATGTTCGGAGCCCGGAAGAGAGACAGCTTCTGATGAGGCCACCAATGGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlyHisIleValMetLeuSerValGly 1679  
 Db 5203 GGGCCACACAGCTTGGCTCCAGATCTGAAACAATCACTCCAGTTTGAAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuProGlyArgAlaAsnAspGlyAspTyr 1699  
 Db 5263 GGGCCCTCCGATGTGAGTCCGTGATGTGTCCGGTGGGTGACCGAGGGGAGTGA 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 Db 5323 CACCACTGTGATCCAGCTGAAGAAATGTTAAGAGAGACAGTGAAGTGAAGCACTGGTGC 5382  
 QY 1717 --LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValAsnLeuGlyProArgLeuHis 1735  
 Db 5383 ACCATGACCTTGAATGTGAGATGAGACACAGAACAGGACATATCGGGGAGCATGCTTCC 5442  
 QY 1736 GlyLeuHisIleLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
 Db 5443 GGGCTAGCGGTAGAGAGCGTGTGTGAGAGCGCTTGAAGACAAAGCTTCCGATGCC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAlaAsn 1775  
 Db 5503 CGTGGATTCGAGAGCTGCATGAGAGAGTGAAGATGGGGGAGCGCCACCAAGCTGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
 Db 5563 ACCCTGAACAATGAACAACGACTCAAGTCAAGGTGAGAGCGCTGTGATGTGAGAGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815

Db 5623 CCTTACCTGAGAGCCCTGTGCCCCCAATAGCCGCTGCACAGAGCTGGAGAGATAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 Db 5683 AGCTGCTGTGACAAAGGATACCTTGGAATTAATCTGTGGATGAGCTGTCACTGAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgIleProSerAlaProHisGlyTyrThrCys 1855  
 Db 5743 CCTCCAGAAACATGGGGGCGCTGCTGCGCTCCCGGCTCCCGAGAGGCTACCTGTGCC 5802  
 QY 1856 GlyCysProProAsnTyrLeuGlyProTyrCysGlyIleThrArgIleAspGlnProCysPro 1875  
 Db 5803 GAGTGTGGCCCACTACCTAGCGGCTGTACTGTGAACAAACTGACCTTCCGTCGCC 5862  
 QY 1876 ArgGlyTyrTrpGlyHisAspThrCysGlyProCysAsnCysAspValSerValGlyPhe 1895  
 Db 5863 AGAGGCTGTGGGGGAACCCGCTGTGTGACCTTGCACCTGTGCGGTCAAGAAAGGCTTT 5922  
 QY 1896 AspProAspCysAsnIleThrSerGlyGlyCysHisCysValGlnAsnHisTyrArgPro 1915  
 Db 5923 GATCCCGACTGTAAATAGACCAACCGCCAGTGCCTCAATGCAGAGAAATTAATCAAGCTC 5982  
 QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 Db 5983 CTAGCCCAAGACCTGTGTGCTGCGACTGTCTCCCAATGGCTCCACAGCCGCACT 6042  
 QY 1936 CysAspProGlnAspGlyGlnCysAspProCysIleProGlyValIleGlyArgGlnCysAsp 1955  
 Db 6043 TGCACATAGGCGACCGGGCAGTGTGCTGCAAGGCCCGGATGATGGCCAGTGCAC 6102  
 QY 1956 ArgCysAspAsnProPheAlaGlyValThrThrAsnGlyCysGlyValAsnTyrAspSer 1975  
 Db 6103 CGCTCGACCAACCCCTTGGCCAGAGTGCACAGCTCGCTGTGAAGTATCTACAGAGTC 6162  
 QY 1976 CysProArgAlaIleGlnAlaGlyIleThrTrpProArgThrArgPheGlyLeuProAla 1995  
 Db 6163 TGTCCAAAGCATTTGAAGCGGCACTGTGTGGCCACACAACTTGGGAGCGGCT 6222  
 QY 1996 AlaAlaProCysProIleGlySerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 Db 6223 GCGGGCCATGCTTAAGGATCCGTTGAATGCGGTCCAGCATGCAAGCGGGAGAG 6282  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuValGly 2035  
 Db 6283 GCTGTGCTGCCCCCAAGCTTTTAATGTACCAACATCTCTTGTGTGACCTCAGGGCC 6342  
 QY 2036 PheAlaGlyArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 Db 6343 ATGAATGAAGACTAGCCCGCATATGACAGCAGTGAACGGGCGCAAGGCTTGAAGCTG 6402  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIle 2075  
 Db 6403 GTAGAGGGCGCTGCGGAGTGTACACAGCAACGGGCAAGCTTGTGGCAATGACGTGCC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 Db 6463 AGGGCTTACAGCTGTGGGCGCACTCTTCAAGCAAGAGCTGGCAGAGGGCTTCAAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisAspThrGlnLeuLeuLeuArgValIleSerAlaLeu 2115  
 Db 6523 CTGGAGCGACAGAGAGCGGCACTTTCAGAGAGAGCTATCACTCGGGAGAGCCCTTC 6582  
 QY 2116 LeuAspThrAlaAsnValArgHisTyrProIleLeuIleGlnIleThrGlnGlyIleThrAla 2135  
 Db 6583 CTGGCCCCAGCACAGGGGCGGCGTGGAGGAGATCCAGCGGAGAGAGGGCGGCAAGGGA 6642  
 QY 2136 TrpLeuLeuGlnHisTyrGlyAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThr 2155  
 Db 6643 CAGCTGTCTGGCGCTGAGGGCTTACTTCAAGAACTGTGAAGCGGATGCTGTGGCGAGAG 6702  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175

Dh 6703 TACCTGGGCGCTTCGTCATCGTCACCGCAACATGATTCTTGCTGTCACATCTTTGAC 6762  
Qy 2176 LysGlyValAsnPheAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGluGlnPro 2195  
Dh 6763 AAGTTCAACTTTTACCGGAGCCAGGGGTCCCGCATTCGACACATCCATGAAAGATTCCCC 6822  
Qy 2196 ProAspLeuGlnLysThrValLysLeuProGluSerAlaPheArg-----Glu 2211  
Dh 6823 AGGAGCTGGAGTCTCTCGTCTCTTCCAGCCGACATTCCTTCAGACCACTGAAAGAAA 6882  
Qy 2212 ThrProValValArgProAlaGly-----ProGly 2222  
Dh 6883 GAAAGCGCCCTGCTGAGGCGCGCTGCGGAGAGACACCCCGACAGACACCGCCGCGG 6942  
Qy 2223 GluAlaGlnLysProGluGluLysAlaArgArgGlnArgArgLysProGluLysSerGln 2242  
Dh 6943 CCTGGCACCGGAGGAGGCGCCGATCAGAGGGGAGGAGCACCTGATGAGCGCTGGC 7002  
Qy 2243 GlyGluAlaValAlaSerValLysLysThrValLeuAlaGlyLeuLeuProLysAsn 2262  
Dh 7003 CAGTTGCGCTCGCTGTCATCATTTACCGACCTGGGGGACGCTCCGCGGAGCGC 7062  
Qy 2263 TyrAspProAspLysArgSerLeuArgValProLysArgProLysLysAsnThrProVal 2282  
Dh 7063 TACGACCCCGACCGTGGCACTCCGTTGCTTCACCGGCGCATCATTAATACCGGATG 7122  
Qy 2283 ValSerLysSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysProVal 2302  
Dh 7123 GTGAGACGCGTGGTTCACGAGGGGGCTCCGCTCCGAGACCCCTGAGAGAGCCGCTC 7182  
Qy 2303 ThrValGlnPheArgLeuGluGlnLysGlnArgThrLysProLysValLysPheTrp 2322  
Dh 7183 CTGGTGAAGTTCCTCGCTGAGGTGAGAGAGAGCAAGCGCTGCTGCGTGTCTGG 7242  
Qy 2323 AsnHisSerLysLeuValSerGlyThrGlyLysTrpSerAlaArgGlyCysGluValVal 2342  
Dh 7243 AACCACTCCCTGGCCGTGGTGGGAGCGGAGGTGGTCTGCCCGGGGCTGAGACTCTG 7302  
Qy 2343 PheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeu 2362  
Dh 7303 TCCAGGAACCGGACATGTCGCTGCGCTGAGTGCACACAGCAGCTTGGCGGTGCTC 7362  
Qy 2363 MetAspValSerArgArgGluAsnGlyGluLysLeuProLeuLysThrLeuThrTrpVal 2382  
Dh 7363 ATGGAATATCTCCAGCGGTGAGAACGGGAGAGTCTGCTCTGAAGATTGTCACCTATGCC 7422  
Qy 2383 AlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeuLeuArg 2402  
Dh 7423 GCTGTGCTCTGTCATCTGGGAGCCCTGCTGCTGCTCTGCTGAGCTGCTGCTGCTG 7482  
Qy 2403 IleLeuArgSerArgGlnHisGlyLysArgArgAsnLeuThrAlaAlaLeuGlyLeuAla 2422  
Dh 7483 ATCTCGCGCTCCAACTGCACACATTCACACAGCACTGCGCTGGGCGCTCTCTCTCT 7542  
Qy 2423 GlnLeuValPheLeuLeuGlyLysLeuGlnAlaAspLeuProPheAlaCysThrValLys 2442  
Dh 7543 CAGCTGGTGTTCGTGATTTGGATCAACAGACGGAACCCGTTTCTGTGTCACAGTGGTT 7602  
Qy 2443 AlaIleLeuLeuHisPheLeuLysLeuCysThrPheSerTrpAlaLeuLeuGlnAlaLeu 2462  
Dh 7603 GCCATCTCTCTCACTACATCTACATGAGCACTTTGCTGAGACCTTCGGAGAGACTG 7662  
Qy 2463 HisLeuLysArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTrp 2482  
Dh 7663 CATGTCTACCGCATGTGACCGAGAGTGCACATTCACACAGGGGCGCATGCGGTCTAC 7722  
Qy 2483 TyrMetLeuGlyLysTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspPro 2502  
Dh 7723 TACGTGTGGGCTGGGAGATCCGCGCATTTGTCAAGAGTGGCGGTGGCTGGAGCCCC 7782  
Qy 2503 GluGlyLysTrpGlyAsnProAspPheCysTrpLeuSerLysTrpAspThrLeuLysTrp 2522  
Dh 7783 CAGGGCTGAGGAGAACCCGACCTTCTGCTGGCTGTGTTCAACACACCTGATTTGGAGC 7842

Qy 2523 PheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuLysLysLeuAlaAla 2542  
Dh 7843 TTGGCGGGGCCCATCGAGACTGTATATCATCAACAGTCACTTCTGCTATCTGCA 7902  
Qy 2543 ArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluLysLysGlyProValSerGlyLeu 2562  
Dh 7903 AAGTTTCTGTGCAAGAAAGACCATTAATTATGGGAAAAAGGATTCGTCTCCGTGCTG 7962  
Qy 2563 GlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSer 2582  
Dh 7963 AGAGCCGATTTCTCTGCTGCTGCTGCTATCAACGCGACCTGGCTGTGGGCTGCTGGCT 8022  
Qy 2583 ValAsnSerAspThrLeuLeuPheHisGlyLysLeuPheAlaThrCysAsnCysLysGlnGly 2602  
Dh 8023 GTGAACCGGAGTCACTGAGCTTTCACTACATCTCTTCGATCTTCAGCGGCTTACAGGGC 8082  
Qy 2603 ProPheIlePheLeuSerTrpValValLeuSerLysGluValArgValAlaLeuLys--- 2621  
Dh 8083 CCTTCGTCCTCTTTCCACTGCGCTGTCACACAGAGGTCCGGAAGCACCTGAAAGGCGC 8142  
Qy 2622 LeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThr---Leu 2640  
Dh 8143 GTGCTGGGGGAGAGAGCTGACCTGAGAGACTCCGCGACACAGAGGCTCCTGCTG 8202  
Qy 2641 ThrSerSerTrpAsnCysProSerProTrpAlaAspGly-----ArgLeuTrpGlnPro 2658  
Dh 8203 AGCGCTCCCTCACTGACCAACACACACTTCGCTGAGAGGGCTGACATGCTGGCACAGAC 8262  
Qy 2659 TyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSer----- 2672  
Dh 8263 TTGGGGAAGTTCACCGCTCGCTGAGACAGATGTCAGGAGTGAAGGATCCAGAGCTC 8322  
Qy 2673 -----GlyLysSerGlnProSerTrpLysPro 2681  
Dh 8323 GCGCTGCTCTGCGGCTGTGAGGGGACCGACCGAGAGACCGAGCGCTCCTCATGCC 8382  
Qy 2682 PheLeuLeuArgGluGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAsp 2701  
Dh 8383 AGAGCTGCAGAGAT-----CCCCCTGGC----- 8406  
Qy 2702 ProGlySerLeuPheLeuGlnGlyGlnArgGlnHisAspProAspThrArgSerAsp 2721  
Dh 8407 -----CACATTTCCACTCAGAT 8424  
Qy 2722 SerAspLeuSerLeuGlnAspAspGlnSerGlySerTrpAlaSerThrHisSerSerAsp 2741  
Dh 8425 AGCAGACTGTCTCTG---GATGAGAGAGCACTTTACGCTCTCACTGTCAGAC 8481  
Qy 2742 SerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2761  
Dh 8482 AGCAGAGACGATGGGAGTGAAGTGAAGAAA-----TGG 8517  
Qy 2762 AspSerLeuLeuGlyProGlyValAlaGlnArgLeuProLeuHisSerThrProLys---Asp 2780  
Dh 8518 GACCGGCGCAAGGGGC-----GCCGTCCACAGACACCCCAAGAGGAGAC 8559  
Qy 2781 GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800  
Dh 8560 GCTGTGGCCAACCAAGTTCGCGCTGCGCTGCGCCGACAGAGCTGGCTGAGAGTCAAGT 8619  
Qy 2801 SerSerGlyAsnGlyAlaPro----- 2807  
Dh 8620 GAGGACCCGAGCGGCAACCCCGCTGAAGTGAACCAAGTCAAGCTGAGCTGAC 8679  
Qy 2808 ---GluGlnArgLeuArgGlnAsnGlyAspAlaLeuSerArgGlnGlySerLeuGlnPro 2826  
Dh 8680 CGCGAGAGACGGGAGTCAACCGTGAAGATACCCCGGACCAAGAGAGCGGGGCGCA 8739  
Qy 2827 LeuProGlySerSerAlaGlnPro-----HisLysGlyLysLeuLysLysCys 2843  
Dh 8740 GCCAGGCTTGTACGACGACGACGCGCCCAAGAGCAAGAGAAAGGATTTGAAAAATTAAGTC 8799





Db 3352 AGTTTCACCCGCGCTGATCGGCTGATCCCGGCCATGACCCCGACGCTGACGACAC 3411  
 QY 1060 LeuThrTyrSerPheGluArgGlyAsnGlnLeuSerLeuValLeuLeuAsnLysSerThr 1079  
 Db 3412 CTCACATCAACCTTCGTCGAGGCGGACAGACTCGGCTGTGTGCTGCTGACCCCGGACAG 3471  
 QY 1080 GlyGlnLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSer 1099  
 Db 3472 GGGGAATCGAGCTGACCGCGGACCTGAGCAACACCGGCGCTGAGGGGCTCATGGAG 3531  
 QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db 3532 GTGTCTGTGTCTGATGAGCATCCACAGCTCCAGCTTCGACCGCTGCGTGCACCATC 3591  
 QY 1120 IleThrAspGlnMetLeuThrHisSerIleThrLeuArgLeuGlnAspMetSerProGln 1139  
 Db 3592 ATCAACGACGACATGCTGACCAACAGCATCTGTCCGCTGGAGAACATGTCCAGGAG 3651  
 QY 1140 ArgPheLeuSerProLeuLeuGlnPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
 Db 3652 AAGTTCCTGTCCCGCTGCTGCGCTCTTCTGTGAGGGGGGTGGCCGCGTGTCTGACCC 3711  
 QY 1160 ProProAspHisAlaValAlaPheAsnValGlnArgAspThrAspAlaProGlyHis 1179  
 Db 3712 ACCAAGACGACGCTCTTCCTTCAACGTCAGAACAGACCCGACGTC--AGCTCCAAC 3768  
 QY 1180 IleLeuAsnValSerLeuSerValGlnProProGlyProGlyGlyProPhe 1199  
 Db 3769 ATCTTAACGTGACCTTTCGCGGCTGCTGCTGCGGCGCTCCGCGC-----CAGTTC 3822  
 QY 1200 LeuProSerGlnAspLeuGlnArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 Db 3823 TTCCCTCTGGAGAGCTGCGAGGACAGATCTAATGAGCGCTGACCAACATC 3882  
 QY 1220 SerAlaGlnArgPheLeuProPheAspAsnAlaCysLeuArgGlnProCysGlnAsn 1239  
 Db 3883 TCACCGACGCGGCTGCTGCTTCGACGACACATCTGCTGCGACGCTCGCGAGAAC 3942  
 QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleLysSer 1259  
 Db 3943 TACATAGAGTGGCTGCTCGTTCGCAATCGACAGCTCGCGCTTCCTTCAGACTCAC 4002  
 QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
 Db 4003 ACCGCTCTCTTCGCGCCATCCACCCCATCAACGCGCTGCTGCGCTCGCGCCGCGC 4062  
 QY 1280 PheThrGlyAspPheCysGlnThrGlnValAspLeuCysPheSerArgProCysGlyPro 1299  
 Db 4063 TTCACGCGGACATGACGAGACGAGATGACCTGCTACTCCAGCCGTCGCGCGCC 4122  
 QY 1300 HisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgProGlyTyrThr 1319  
 Db 4123 AACGGCGGCTGCGCGACGCGCGAGGGGCTACACCTGCGAGTGTTCAGAGACTCACT 4182  
 QY 1320 GlyGlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysLysAsn 1339  
 Db 4183 GAGAGACATGTGAGGTGAGTGCCTCGCTGACGCGCTGTGCCAACGCGGTGTCCAGAAC 4242  
 QY 1340 GlyGlyThrCysValAsnLeuValGlyGlyPheLysCysAspCysProSerArgLysAsp 1359  
 Db 4243 GGGGGGACCTGCGGTAACCTGCTCATCGGCGGCTTCACATGCTGTCTCTCGGAGAG 4302  
 QY 1360 PheGlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
 Db 4303 TATGAGAGGCGCTACTGTGAGGTGACACACAGAGGCTTCGCGCCAGTCTTCGTGACC 4362  
 QY 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGln 1399  
 Db 4363 TTCGGGGGCTGAGACGCGCTTCCTCACTCAATCCCTCCCTCAAGTTTGCACTCGAGAA 4422  
 QY 1400 ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlnLysHisAspPheValAlaLeu 1419

Db 4423 AGAAAGGCTGCTCTCTCAACACGCGCTTCAATGAGAACACGACTTCATCGCCCTG 4482  
 QY 1420 GlnValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlyLysThrThrVal 1439  
 Db 4483 GAGATCGTGGACGACAGGTGTGACTTCTTCGACGGGAGAGACAAACGACCGTG 4542  
 QY 1440 SerProPheValProGlyValSerAspGlyGlnThrPheIleThrValGlnLeuLysTyr 1459  
 Db 4543 GCACCGAAGTTCACAGTGTGTGAGTACGCGGCGGTGGCACTGTGTGACGTTGACGATC 4602  
 QY 1460 TyrAsnLysProLeuLeuGlnGlnThrGlyLeuProGlnGlyProSerGlnLysVal 1479  
 Db 4603 TCAACAAAGCCCAATATTTGGCACTGGAGCTGCCCCCATGGGCGCTCCGGGAAAAGATG 4662  
 QY 1480 AlaValAlaThrValAspGlyCysAspThrGlyValAlaLeuAspPheGlySerValLeu 1499  
 Db 4663 GCGGTGTGACAGTGTGATGTGTACACACCAATGCTGTGTGCGCTTTGGAAAGACATC 4722  
 QY 1500 GlyAsnTyrSerCysAlaAlaGlnGlnThrGlnGlySerLysLysSerLeuAspLeu 1519  
 Db 4723 GGGAACTACAGTGGCGCTGCCAGGGCACTCAGACCGGCTCCAAAGAAAGTCCCTGATCTG 4782  
 QY 1520 ThrGlyProLeuLeuLeuGlnGlyValProAspLeuProGlnLysPheProValArgMet 1539  
 Db 4783 ACCGCGCTCTACTCTCGGGGGGTGTCCCACTGCGCAAGAAAGCTCCAGTGCACAAAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
 Db 4843 CGGCACTGTGTGGGTGTGATGAGAACCTGTCACTGATCGAACGAAATGTGACATGGCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAsp 1579  
 Db 4903 GGATTCATCGCCCAAAATGGACCCGGAAGGCTCGCTGCGAGAGAACTTCGCGCAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrPheAspAlaPheSerCysGln 1599  
 Db 4963 GGGAGCGGCTGTCAAAATGAGGACCTGTGTCAACAGGTGAATATGTATGTGTGAG 5022  
 QY 1600 CysProLeuGlnPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPhe 1619  
 Db 5023 TGTCCACTCCGATTCGGCGGAGAAATGTGAGCAAGCATCCCAACCCGACGCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaThrPheIleGlyLeuSerLeuProIleSerGlnProTyr 1639  
 Db 5083 ACGGCTGAGAGGTGTGTCTCGAGTGCATGAACTATCATCTGTGTGCTCGGTATC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTGGGCTCATGTTCCGACCCGGAAGAGACAGAGCTTCTATGAGGCGCACAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGly 1679  
 Db 5203 GGGCCACACAGCTTTCGCTCCAGATCCGAAACAATCACTCCATGTAAGGTGTCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspThr 1699  
 Db 5263 GGCCTCTCGATGTGAGTGCCTGATGTCTCCGGGTGTGCGGGTGAACCAAGGAGATG 5322  
 QY 1700 HisIleAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 Db 5323 CACCACTGCTGATGAGTGAAGAATGTAAAGAGACAGTGAAGATGAAGACCTGTGTC 5382  
 QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
 Db 5383 ACCATGACCTGATGATGTGAGTGAACCAAGAACAGGACAGATGTGGGGGCAATGCTCCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
 Db 5443 GGGCTGACGCTAAGAGAGGTGTGTGTGAGGCGCTCTGAAGACAGAGTCTTCGAGCC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlnValArgValSerAspThrProGlnGlyValAsn 1775  
 Db 5503 CGTGAATTCGAGGCTGATGACAGGAGTGAAGATGGGGGGGACCCCAACAGTCCGCC 5562



QY	1776	SettleuapProserNieslgyluserl1easvna1glunglgyCysserleuproaer	1795
Dp	5563	ACCTGAACAAAGAACAGCACTCAAGGTCAAGGTGAAGAGACGGCTGTGAATGTGACAC	5622
QY	1796	ProCyaaerSerAenProCyProAlaambSerTyCysserAanbPrtPaerSerTy	1815
Dp	5623	CCCTTAACCTGAAGCCCTGTCTCCCAATAGCCGCTCCACAGACGCTGGAGAGATAC	5682
QY	1816	SerCysserCyaaerProGlyTyTrgTyaaerAenCyvthAanValCyaaerLeuAen	1835
Dp	5683	AGCTGCGCTGTGACAAAGGATACCTTGAAATTAACCTGTGAATGCTGTCACTGAC	5742
QY	1836	ProCyssgynH1eGlnserValCyvthAargyuvProserAlaProH1eGlyTyTrthCyv	1855
Dp	5743	CCCTGCGAAGAACAGGGGGCGCTGGCTGGCTCCCGGCTCCCGACAGGCTAGTGTGC	5802
QY	1856	GluCyvProProAanTyLeuGlnTyProTyCyvsglnThAargl1eAerGlnProCyvPro	1875
Dp	5803	GAGTGTGGGCGCCAGTCACTACGGGCGGTACTGTGAAGAACAACTGACCTTCCTGCCCC	5862
QY	1876	ArgGlyTrpTrgTylnH1eProThCyvsgTyProCyvAenCyaaerValSerTyeglyPhe	1895
Dp	5863	AGAGGCTGTGTGGGAGAACCCGCTGCTGTGACCCCTGCACTGTCGCTCAGCAAGGCTTT	5922
QY	1896	AspProaerCyvAanTyvThSerGlyGluCyvH1eCyvlyvGlnAenH1eTyAargPro	1915
Dp	5923	GATCCCACTGTATTAAGACCAAGCCAGTGCACAAAGCAAGAAATTAACAACCTC	5982
QY	1916	ProGlySerProThH1eCyvLeuLeuCyvAerCyvTyTrProThGlySerLeuSerArgVal	1935
Dp	5983	CTACCCCAAGACACTGTCTGCTGCGACTGCACTGCTCCCAATGGCTCCCAAGCCGACT	6042
QY	1936	CyaaerProGluAerGlyGlnCyvProCyvLyvProGlyVal11eGlyAargGlnCyvAer	1955
Dp	6043	TGCACACATGGCCACCGGCGAGTGTGCTCGCAAGCCCGGCTCATCGCGCCCACTGCAAC	6102
QY	1956	ArgCyvAerAenProPheAlaGlnValThThAanGlyCyvsgValAanTyAarSer	1975
Dp	6103	CGCTGCGACACCGGTTTGCGAAGGTACACAGCTCGGCTGTGAAGTATCTAACAGGCG	6162
QY	1976	CyAerProAargAla11eGlnAagly11eTrpTrProAargThAArgPheGlyLeuProAla	1995
Dp	6163	TGTCCCAAGCAATTGAAGCGCGGACTGTGTGGGCGACAGACCAAGTTTGCGACCCGGCT	6222
QY	1996	AlaAlaProCyvProLyvGlySerPheGlyTyThAlaValAargH1eCyvAerGlnH1eAarg	2015
Dp	6223	GCGGTGCAATGCCCTTAAGGATCCGTGTGAATCGGTCCACACTGACGCGGAGAACAG	6282
QY	2016	GlyTyTrleuProProAanTyPheAenCyvthSer11eThPheAenGlnTyLeuGly	2035
Dp	6283	GCGCTGCTGCCCCAGACGCTTTTAATCTGACACCAACTCTCTTCGTGAACCTCAAGGCG	6342
QY	2036	PheAlaGlnAargLeuGlnAargAenGlySerGlyLeuAerSerGlyAargSerGlnGlnLeu	2055
Dp	6343	ATGAATAGAAGAGCTGAGCGCGCAATGAGCGAGGTGAGCGCGCGCCCTGCACTG	6402
QY	2056	AlaLeuLeuLeuAargAenAlaThGlnH1eThThAlaGlyTyTrPheGlySerAerValLyv	2075
Dp	6403	GTTGAGGCGCTGCGACAGTCAACAGACACGCGGCAOGCTTTTGCAATGAGTGTGCG	6462
QY	2076	ValAlaTyGlnLeuAlaThArgLeuLeuAla11eGlnuserThGlnAargGlyPheGly	2095
Dp	6463	ACGCGCTTAACAGCTGTGCGGCAAGTCTTTACAGACACAGAGCTGGACAGAGGCTTGAC	6522
QY	2096	LeuSerAlaThGlnAerValH1ePheThTrGlnAenLeuAerValGlySerAlaLeu	2115
Dp	6523	CTGGCAGCGCAACGACGACGCGCACTTTTCAAGAGCGTCAACACTCGGAGAGGCGCTC	6582
QY	2116	LeuAerPhH1eAaenLyvAargH1eTrpGlnLeu11eGlnGlnThTrGlnGlyTyThAla	2135
Dp	6583	CTGGCCCCAGCACAGGCGGCGGTGGAGAGCATTCAGAGGAGACGAGGCGGCGACAGCA	6642

QY	2136	TrpleuLeuEnglnIhIeTYrGluAlaTYrAlaSerAlaLeuAlaGlnAsnMetArgIhIeThr	2155
Db	6643	CAGCTGCTCCGGCCGCTCGAGGGCTACTTCAGCAACCGGAGCGACAGGATGGTCGGGGAGCG	6702
QY	2156	TYrLeuSerProPheThrIleValThrProAsnIleValIleSerValIValArgLeuSp	2175
Db	6703	TACCTGGGGCCCTTCGTGATGTGTACCCGCCAACATGATTCCTTGGCTGTGACATCTTAC	6762
QY	2176	LySGlyAsnPheAlaGlyAlaLysLeuProArgTYrGluAlaLeuArgGlyGluGlnPro	2195
Db	6763	AAGTTCACTTACGGAGCGACGGAGTCCGGGATTCGACACATCCATCAAGAAAGATTCGCC	6822
QY	2196	ProAsnLeuGlnThrValIleLeuProIuSerValPheArg-----Glu	2211
Db	6823	AGGAGAGCTGGAGTCTCCGTCTCTCCACGCCACTTCCTACACACTGAAAGAAAA	6882
QY	2212	ThrProValValArgProAlaGly-----ProGly	2222
Db	6883	GAAAGGCCCTCTGAGGGCCGGCTGGCCCGAGAGCAACCCCGAGACCAACGCGCCGGGG	6942
QY	2223	GluAlaGlnGluProGluGluLeuAlaArgArgGlnArgAlaGlnProGluLeuSerGln	2242
Db	6943	CTGGGCACCGAGAGAGAGGCCCGCATCAGCAGGGAGGAGCAGACCTCGAAGAGCTGGC	7002
QY	2243	GlyGluAlaValAlaSerValIleIeTYrArgThrLeuAlaGlyLeuLeuProIhIeAsn	2262
Db	7003	CAGTTCCGCCGTCCGTCTGTGTATCATTTACCGCACCTCGGGGAGCTCTCCCGAGGCC	7062
QY	2263	TYrAspProAspLysValArgSerLeuArgValProLysArgProIleIeAsnThrProVal	2282
Db	7063	TACACACCCCGACCGTCCAGCTCCGGGTGCTCACCGGCCCACTTAATAAACCCGATG	7122
QY	2283	ValSerIleSerValIhIeAspAspGluGluLeuLeuProArgAlaLeuAspLysProVal	2302
Db	7123	CTAGACACCGCTGGTGTACAGCGAGGGGGCTCCGCTCCGAGACCCCTGGAGAGGCCGTC	7182
QY	2303	ThrValGlnPheArgLeuLeuGlnThrGluGluArgThrLysProIleCysValPheTyr	2322
Db	7183	CTGGTGGAGTTCGCCCTGCTGAGAGTGAAGGAGGAAACCAAGCTGTCTGGTGTCTGG	7242
QY	2323	AsnIhIeSerIleLeuValSerGlyThrGlyGlyTYrSerIleAlaArgGlyCysGluValAl	2342
Db	7243	AACCACTCCCTGAGCCGTGGTGGGAGCGAGAGGTGTCTGCCCCGGGCTGGAGACTCTG	7302
QY	2343	PheArgAsnGluSerIhIeValSerCysGlyGlnCysAsnIhIeMetThrSerPheAlaValLeu	2362
Db	7303	TTCAAGAACCGGACACATGTGCTCTGCACGTGAGCCACACAGCACACTTTCGGGTGCTC	7362
QY	2363	MetAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThrTYrVal	2382
Db	7363	ATGGAATATCTCCAGGCGCTGAGAAAGGGGAGAGTCTGCTCTGAAAGATTGCACTTAATGCC	7422
QY	2383	AlaLeuGluValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArg	2402
Db	7423	GCCTGTGCTCTGTACATGGGAGCCCTCTGTGTGCTTCCTGCTCTGAGCTGTGCTCCG	7482
QY	2403	IleLeuArgSerAsnGlnIhIeGlyTYrLeaArgAsnLeuThrAlaAlaLeuGlyLeuAla	2422
Db	7483	ATGCTGGCGCTCCACCTGCAACGACTTCAACAGCACTCCCGCGGGCGCTTCCTCTCTCT	7542
QY	2423	GlnLeuValPheLeuLeuGlyYIleAsnGlnAlaAspLeuProPheAlaCysThrValIle	2442
Db	7543	CAGCTGTGTTCTGTGATGGAGATCAACAGGAAACCGGTTCGTGTGACAGTGGATT	7602
QY	2443	AlaIleLeuLeuIhIePheLeuTYrLeuCysThrPheSerTyrAlaLeuLeuGluAlaLeu	2462
Db	7603	GCCATCTCTCTCCACTCACTCACTCATAGACACCTTTCCTGAGACCTCGTGGAGAGCTG	7662
QY	2463	HISLeuTYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTYr	2482
Db	7663	CATGTCTTACCGCATGCTGACCGAGGTGGCAACATCAACCGGGGCCATGGCGGTCTTAC	7722
QY	2483	TYrIleSerLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspPro	2502





Db 2272 TACGTGCGGCGTGAACGATCCGACGCGGTGCGACATCGCGCATGCTTAATC 2331  
QY AsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisArgThrValAsn 720  
Db 2332 AACGTCACTATGCTCAACACCCAGCGCTGTCTTTCAGAGCTCCCTCAACACGAGT 2391  
QY ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
Db 2392 GTACAGTGAAGACAGCGCTGTGGGCACTTCATGCTATCCCTCAAGTGCACATGAGAC 2451  
QY ThrGlyGluAsnAlaArgGlyLeuThrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACAGGAGAGATGCGCGCATCACTAGTATTCAGAGATCCCGCGCATTCGCGCATC 2511  
QY AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspArgGluAspGlnValSer 780  
Db 2512 GACCCGACAGTGGCGACCATGATACATGATGAGCTGAGCTATGAGAACAGAGTGGCC 2571  
QY TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db 2572 TACACGCTGACATCAATGCGCCAGAGCAACGCGATCCCGCAAAATCAGACACACAC 2631  
QY LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db 2632 CTAGAGATCTCATCTCGATGCGCAATGACATGCAACCCAGTCTCTGTGGATTTCTAC 2691  
QY GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGGTTCATCTGTGAGAGTCTCACCTCGACGACATCTCCAGGCTCTGCGCACG 2751  
QY AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly 860  
Db 2752 GACCGGAGCTCAGAGTCCCAATGCGCGCTCTGTAACCTTCCAGGGTGGGAGCAGACGG 2811  
QY AspGlyAspPheIleValGluSerThrSerGlyIleValAlaArgThrLeuArgGluAsp 880  
Db 2812 GATGGGACTCTTCAACGAGCCGACGCTCGGTGTGATTCGACCCAGCCCGCTGGAGC 2871  
QY ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspArgIleMetPro 899  
Db 2872 CGGAGAGATTTGGCCGCTGACACCTTTGGGCTCTGCTGTGATCCGGGCAATCCAC 2931  
QY ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
Db 2932 CCCCTTAGCGCTCGGTGAAGAAATCCAGGTGACCATTTGACATTAATGACATGCCCC 2991  
QY ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTGTGAGAGAGACGAACTGAGCTGTGTGTGAGAGAGAACCAACCGAGGGGTGGTG 3051  
QY ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGCAAGATTTGCTGTACGACCTGATGAGGCCCTTAATGCCAGATCATGTATCAG 3111  
QY IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyLeuLeuThr 979  
Db 3112 ATTGTGAAAGGACATGCGGCATTTCTTCACTGACCTGACCTGCAACGGGAGCTGCGT 3171  
QY AlaLeuValAspLeuAspArgTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
Db 3172 GCCATGTGAGCTGAGCTTTGAGGTCGGCGGAGAGATGTGCTGTGGTGGCGAGCCACG 3231  
QY SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCGCTGTGGAGCCGAGCCAGCTGCATCTTCTCGTGGACCGAATGACAC 3291  
QY ProProValLeuGluAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3292 CCGCTGTGTGCTGCGCGCATTCGATCTCTCAACAATATGTCACCAACAGTCCAC 3351  
QY SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGCGGTGATGGCTGATCCGCGCCATACCCCGAGGTGACAGC 3411

QY LeuThrTyrSerPheGluArgIleAsnGluLeuSerLeuValIleuLeuAsnAlaSerThr 1060  
Db 3412 CTCACCTACACCTTGTGTGAGGAGCAACAGCTGCGCTTGTCTGTGACACCCCGCACG 3471  
QY GlyLeuLeuValLeuSerArgAlaLeuAspAsnAspArgProLeuGluAlaIleMetSer 1099  
Db 3472 GCGAATCTGACCTAGCGCGGACCTGAGCAACAACCGCGCTGGAGCGCTCATGAG 3531  
QY ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 3532 GTGTCTGTCTGATGGATTCACAGCGCTCAGCGCTTCTGACCCCTGTGTCTACATC 3591  
QY IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Db 3592 ATCAGGAGACATGCTGACCAACAGATCACTGCTCCGCTGGAACATGTCCAGAG 3651  
QY ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 3652 AGTTCTGTCTCCCGCTGTGGCCCTTCTGTGAGAGGGGTGGCGCGGTGCTGCAC 3711  
QY ProProAspHisValValPheAsnValGlnAlaGlyThrAspAlaProGlyGlyHis 1179  
Db 3712 ACCAAGAGACGATCTTGTCTTCAACGTCGAAACGACACGACGATC 3768  
QY IleLeuAsnValSerLeuSerValGlyGlnProProGlyGlyIleProProPhe 1199  
Db 3768 ATCTGAACTACCTTCTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3822  
QY LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
Db 3823 TTCCTGTGAGAGACTGAGAGAGAGATCTTCAATCGAACCTGTGTGACACATC 3882  
QY SerAlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Db 3883 TCCAGCAGCGGTGCTGCTCTTCAACACATCTGCTGCGGAGGCTGCGAGAAC 3942  
QY TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 3943 TACATGAGTGTGTTCGCTTGTGATTCGACATCTCGCGCGCTCTCTCAGCTCAC 4002  
QY SerValLeuPheArgProIleHisProValGlyLeuArgCysArgCysProProGly 1279  
Db 4003 ACCGTCTCTTCCCGCCCATCCACCCCATCAACGCGCTGCGCGCGCGCGCGCGCG 4062  
QY PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 4063 TTCACCGGAGTACTGCGAGAGAGAGATCGACTCTGCTACTCCGACCGGTGCGCGC 4122  
QY HisGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 4123 AACGCGCGCTGCGGAGCGGAGCGGAGGCTGACATCTGAGATGCTTGGAGACTTCACT 4182  
QY GlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysValAsn 1339  
Db 4183 GAGAGGACTGTGAGGTGATGCCGCTCAGCGCCCTGTGCCAACGGGGTGTGCAAGAAC 4242  
QY GlyGlyThrCysValAsnLeuLeuValGlyGlyPheCysAspCysProSerGlyAsp 1359  
Db 4243 GGGGACCTGTGAGACCTGCTCATTCGCGGCTTCCACTGCTGTGCTCTCTGCGAG 4302  
QY PheGluValProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Db 4303 TATGAGAGGCTTACTGTAGAGTGAACACAGAGACTTCCGCGCCAGCTTCTGTCAAC 4362  
QY PheArgGlyLeuArgAlaArgPheHisPheThrLeuAlaLeuSerPheAlaThrArgGlu 1399  
Db 4363 TTCGGGGCTTGAACAGGCTTCACTTCAACATCTTCTCAAGTTTGTCCACTCAGAA 4422  
QY ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluValHisAspPheValAlaLeu 1419  
Db 4423 AGGAACGCTTCTCTTCAACAGCGCGCTTCAATGAGAACGACATTCATCCCGCTG 4482

QY 1420 GlnValIleGlnIleGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrThrVal 1439  
 Db 4483 GAAATCTGACGACGAGCTGACGCTCACTTCTCTGACGCGACAAACACGACGCTG 4542  
 QY 1440 SerProPheValProGlyGlyValSerAspGlyGlnTrpHisThrValGlnLeuValTyr 1459  
 Db 4543 GCACCGAAGGTTCCCACTGGTGTGATGACGGGGGGGACCTCTGTCGACGGGTACTAC 4602  
 QY 1460 TyrAsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnVal 1479  
 Db 4603 TACACAAAGCCCAATATTGGCACCTGGGCTGCCCCATGGGCGCTCGGGAAGATG 4662  
 QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 Db 4663 GCGGTGTGACAGTGAATGTGACACAACTGCTGCGCTTGGAAAGACATC 4722  
 QY 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerValSerLeuAspLeu 1519  
 Db 4723 GGGAACTACAGCTGCGCTGCCCCAGGACCTACAGCCGCTCCAAAGATCCCTGGATCTG 4782  
 QY 1520 ThrGlyProLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMet 1539  
 Db 4783 ACCGGCCCTTACTCTCGGGGGGTGTCCCACTGTCAGAAAGATTCCCACTGTCACAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
 Db 4843 CCGCAGTTCGTGGCTGTCACGCGAACCTGTCACTCCAGCAAAATGTGGACATGGCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaValSerValCysAsp 1579  
 Db 4903 GGAATTCATCCCAACATATGGACCCGGAAAGCTGCGCTCTCGAAGAACTTCTGGAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAspAlaPheSerCysGlu 1599  
 Db 4963 GGGAGCGGTGTGACGATGAGGACCTGTGTCAACAGTGAATATGATCTGTGTGAG 5022  
 QY 1600 CysProLeuGlyPheGlyGlyValSerCysAlaGlnIleMetAlaAsnProGlnHisPhe 1619  
 Db 5023 TGTTCATCTCGAATTCGCGGGAAGAACTGTGACAGCAAGCCATGCTCCACCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyr 1639  
 Db 5083 AGCGGTGAGAGCTCGTGTCTGTGACGACCTTAACATCATCTCTGTGCTGTGATC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTGGGGCTCATGTTCCGAGCCCGGAAGAGACAGAGCTTGTATGAGGACCAAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlyHisIleValMetLeuSerValGlyGly 1679  
 Db 5203 GGGCCCAACGCTTTCGCTCCGATCTGTACAACTAAGCTTTCAGTTGAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTrp 1699  
 Db 5263 GGGCCCTCCGATGTGAGTCCGATGCTGTCCGGGTGCGGGTACCCGAGGAGATG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 Db 5323 CACCACTGCTGATCGAGTGAAGAAATGTTAAGAGACAGTGAAGACCTGCTG 5382  
 QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValLeuGlyProArgLeuHis 1735  
 Db 5383 ACCATGACTTGTGACTATGGGATGACACAGAACAGAGATCCGGGGGCATGCTTCCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
 Db 5443 GGGGTACGATAGAGACGCTGTGTGAGGCGCTCTGAAGACAAAGTCTCCGTGCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspTrpProGlyGlyValAsn 1775  
 Db 5503 CGTGGATTCCGAGGCTGACGAGGAGTGAAGATGGGGGGACGCGCCACCAACGTCGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795

Db 5563 ACCCTGAACATGAACAACGCACTCAAGGTGAGAGGACGGCTGTATGTGACGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyr 1815  
 Db 5623 CCTGTACTCGAAGCCCTGTCTCCCAATAGCGCGCTGCAACAACCTGGAGGACTAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 Db 5683 AGCTGGCTGTGTGACAAAGGTACTTGTGAATTAATCTGTGTGATGCTGTACTGAAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgLysProSerAlaProHisGlyTyrThrCys 1855  
 Db 5743 CCTGGAGAACATGGGGGCGCTGCGAGCGCTCCCGGCTCCCGCAGGGCTACGTGTGC 5802  
 QY 1856 GlnCysProProAsnTyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
 Db 5803 GAGTGTGGCCCACTACTACGAGCGGTACTGTGAACAAACTGCATCCCTCGTCCGCC 5862  
 QY 1876 ArgGlyTyrTrpGlyHisProThrCysGlyProCysAsnCysAspValSerLysGlyPhe 1895  
 Db 5863 AAGAGCTGTGGGGGAACCCCTGTGTGACCTGTGCATCTGTGCGCTCAGCAAGGCTTT 5922  
 QY 1896 AspProAspCysAsnIleThrSerGlyGlnCysHisCysValGlnAsnHisTyrArgPro 1915  
 Db 5923 GATCCCACTGTATTAAGACCAACGGCCAGTGCACATGACAGAGAAATTACTACAGCTC 5982  
 QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 Db 5983 CTAGCCCAAGACACCTGTGTCTGTGACCTGTGACCTTCCCAATGCGCTCCACGCGCACT 6042  
 QY 1936 CysAspProGlnAspGlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAsp 1955  
 Db 6043 TGGACATGGCCACCGGGAGAGTGTGCTGCACACCCGGGTGATCGCGCCGACATCAAC 6102  
 QY 1956 ArgCysAspAsnProPheAlaGlnValThrThrAsnGlyCysGlnValAsnTyrAspSer 1975  
 Db 6103 CCTGTGACAAACCGTTGTCCAGAGTACCAAGCTCGCGCTGTGAAGTGTACTACATGTC 6162  
 QY 1976 CysProArgAlaIleGlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAla 1995  
 Db 6163 TGTCCAAAGCAATTGAGCGCGCATCTGTGGCCCAACAAAGTTCCGGACCGGCT 6222  
 QY 1996 AlaAlaProCysProLysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 Db 6223 GCGGTGCAATGCTTAAGAGATCCGTGGAATGCGGTCCGACACTGCAACGGGAGAG 6282  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlnLeuLysGly 2035  
 Db 6283 GCTGGCTGCCCCAGAGCTTTTAATGTATCAACCATCTCTGTGTGACCTCAGGGCC 6342  
 QY 2036 PheAlaGlnArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 Db 6343 ATGAATGAAGACTAGCCGCAATAGACGAGGTGAAGCGCCAGCGCCCTTGACGTG 6402  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLys 2075  
 Db 6403 GTGAGGGCGCTGCGAGTCTACAGACAGACAGGGCACGCTTTGTGCAATACGTGCC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 Db 6463 ACGGCTTACAGCTCTGGGCCACGCTCTTACAGACAGAGCTGGCAGAGGGCTTGCAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlyAsnLeuLeuValGlySerLeuLeu 2115  
 Db 6523 CTGGAGCCCAAGAGAGCGGCACTTTACAGAGAGGTATCACTCGGAGCGCGCTC 6582  
 QY 2116 LeuAspThrAlaAsnLysArgHisTyrGlyLeuLeuGlnGlnThrGlnGlyGlyVal 2135  
 Db 6583 CTGGGCCCAAGCAACGAGCGCGCTGGAGAGATCCAGCGGAGAGGGGCGGCAACGGCA 6642  
 QY 2136 TrpLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAspMetArgHisThr 2155

Db 6643 CAGCTGCTCCGGGCTCGAGGGCTACTTCAGACGTCGACGCAACCTGCGCGAGCG 6702  
 Qy 2156 TyLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAap 2175  
 Db 6703 TACCTCGGGCCCTTCCTGATCGTCACCGCCACATGATTCCTGCTCCACATCTTGAC 6762  
 Qy 2176 TyGGIYAsnPheAlaGlyAlaValLeuProArgTyGluAlaLeuArgIlyGluInPro 2195  
 Db 6763 AACCTTCAACTTTCAGGAGGACGAGGTCCTCCGATTCGACACATCATGAGAGTTCCTCC 6822  
 Qy 2196 ProAspLeuGluThrValIleLeuProGluSerValPheArg-----Glu 2211  
 Db 6823 AGGAGCTGAGTCTCCCTGCTCTTCCACCGACTTCTTCAGACCACCTGAAGAAAAA 6882  
 Qy 2212 ThrProValValArgProAlaGly-----ProGly 2222  
 Db 6883 GAGAGCCCTGCTGAGGGCCGGCTGGCCGAGAGCACCCCGGACACACCGCCCGGG 6942  
 Qy 2223 GluAlaGluInuProGluLeuAlaArgArgIlnArgIlnArgIlnArgIlnArgIln 2242  
 Db 6943 CCTGGACCGGAGGAGGAGGCGCCGATCAGAGGCGGAGGCGACACCTGATGACGCTGGC 7002  
 Qy 2243 GlyValAlaValAlaSerValIleIleTyArgThrLeuAlaGlyLeuProHisAan 2262  
 Db 7003 CAGTTCGCGCTGCTGCTGATCATATTCACGACCCCTGGGCGACTCTGCGCCGAGCCG 7062  
 Qy 2263 TyAspProAspIlyAsnSerLeuArgValProIlyArgProIleIleAsnThrProVal 2282  
 Db 7063 TACGACCCGACCGGACGCTCCGGTTCCTCACCGGCCATCATTAATACCCGAGT 7122  
 Qy 2283 ValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspIlyProVal 2302  
 Db 7123 GTGAGACCGCTGGTGTACAGCGAGGGGCTCCGCTCCGAGACCCCTGGAGAGGCCGCTC 7182  
 Qy 2303 ThrValGluPheArgLeuLeuGluInuArgIlnArgIlnArgIlnArgIlnArgIln 2322  
 Db 7183 CTGCTGAGATTCCTCCCTGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7242  
 Qy 2323 AsnHisSerIleLeuValSerGlyTyArgIlyTyArgIlyTyArgIlyTyArgIlyTyArg 2342  
 Db 7243 AACCACTCCCTGCGCTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7302  
 Qy 2343 PheArgAsnGluSerHisValSerGlyGlnGlyAsnHisMetThrSerPheAlaValLeu 2362  
 Db 7303 TCCAGAGACCGGACACATGTCCTGCTGACATGACACACACAGCTTGGGCTGCTC 7362  
 Qy 2363 MetAspValSerArgArgIlnArgIlnArgIlnArgIlnArgIlnArgIlnArgIlnArg 2382  
 Db 7363 ATGGATATCTCCAGGCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7422  
 Qy 2383 AlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeuLeuArg 2402  
 Db 7423 GCTGTGCTCTTCATCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7482  
 Qy 2403 IleLeuArgSerArgIlnArgIlnArgIlnArgIlnArgIlnArgIlnArgIlnArg 2422  
 Db 7483 ATGCTCGCTCCCAACCTGACACATTCACACACACCTCCGCTGGCGCTCTCTCTCTCT 7542  
 Qy 2423 GluLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIle 2442  
 Db 7543 CAGCTGCTGCTGCTGATGGATTCACACACACGAAAAACCTGTTCTGTCACAGTGGT 7602  
 Qy 2443 AlaIleLeuLeuHisPheLeuTyLeuCysThrPheSerTrpAlaLeuLeuGlnAlaLeu 2462  
 Db 7603 GCCATCTCTCTCACTACATCTACATGACACCTTCTGACACCTCGGAGAGAGCTCG 7662  
 Qy 2463 HisLeuTyArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTy 2482  
 Db 7663 CATGTCTACCGACATGCTGACGAGGTGACATTCACACGAGGCGCCATGCGGTCTAC 7722  
 Qy 2483 TyMetLeuGlyTyArgIlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAapPro 2502  
 Db 7723 TACGTGTGGGCTGGGGCATCCCGGCATTTGTCTACAGAGACTGGCGGTGCGCTGAGACCC 7782

Qy 2503 GluGlyTyArgIlyAsnProAspPheCysTrpLeuSerIleTyArgThrLeuIleTrpSer 2522  
 Db 7783 CAGGGCTACGGGAACCCGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7842  
 Qy 2523 PheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyIleLeuAlaAla 2542  
 Db 7843 TTTGGGGGGCCCATCGAGCTGTATATCATCAACACAGTCACTTGTCTCTATCTGCA 7902  
 Qy 2543 ArgAlaSerCysAlaAlaGluArgIlnArgIlnArgIlnArgIlnArgIlnArgIlnArg 2562  
 Db 7903 AAGTTTCTGCAAAAGAAAGACATTAATTAAGGAAAAAGGATGCTCTCCCTGCTG 7962  
 Qy 2563 GlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuAlaLeuLeuSer 2582  
 Db 7963 AGGACCGGATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8022  
 Qy 2583 ValAsnSerAspThrLeuLeuPheHisTyIleLeuPheAlaThrCysAsnCysIleGlnGly 2602  
 Db 8023 GTGAACCGGATGCACTGATGCTTCACTACCTCTTCCGCTTCCTTACAGGCGCTTACAG 8082  
 Qy 2603 ProPheIlePheLeuSerTyValValLeuSerIlyGluValArgIlyAlaLeuLeuSer 2621  
 Db 8083 CCTTGTCTCTCTTCTTTCACCTGCTGCTCAACAGAGGTCCGAGAGCACTTAAGGCGC 8142  
 Qy 2622 LeuAlaCysSerArgIlyProSerProAspProAlaLeuThrThrIlySerThr---Leu 2640  
 Db 8143 GTGCTCGGCGGAGGAGACCTGACCTGAGAGACTCCGACACACAGGAGGACACCTGCTG 8202  
 Qy 2641 ThrSerSerTyArgCysProSerProTyAlaAspGly-----ArgLeuTyGlnPro 2658  
 Db 8203 AGCGCTCTCTCACTGACCAACCACTTCGCTGAGCGGCGCTGACATGCTGCGCACAGAC 8262  
 Qy 2659 TyGlyAspSerAlaGlySerLeuHisSerThrSerArgSer----- 2672  
 Db 8263 TTGGGAGATCCACCGCTCGCTGACAGATCGTCAGAGATGAAGGATCCAGAACTC 8322  
 Qy 2673 -----GlyIlySerGlnProSerTyIlePro 2681  
 Db 8323 GCGGTGCTCTGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8382  
 Qy 2682 PheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAsp 2701  
 Db 8383 AGAGCTCAAGAT-----CCCCCTGGC----- 8406  
 Qy 2702 ProGlySerLeuPheLeuGluGlyGluAspGlnGlnHisAspProAspThrAspSerAsp 2721  
 Db 8407 -----CACGATTCGATCTCAAT 8424  
 Qy 2722 SerAspLeuSerLeuGluAspAspGlnSerGlySerTyAlaSerThrHisSerSerAsp 2741  
 Db 8425 AGCGAGCTGTCCCTG---GATGAGCAGAGCAGCTCTTAAGCTCTCTCACTGCTCAAGC 8481  
 Qy 2742 SerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2761  
 Db 8482 AGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8517  
 Qy 2762 AspSerLeuLeuGlyProGlyValAlaGluArgLeuProLeuHisSerThrProIlyS---Asp 2780  
 Db 8518 GACCCGCGCAAGGCG-----GCCGTCCACAGACCCCCCAAGAGGAGC 8559  
 Qy 2781 GlyIlyProGlyProGlyValAlaProTrpProGlyAspPheGlyThrAlaIlyGlu 2800  
 Db 8560 GCTGTGGCCACACAGCTTCGCGCGCGCTGCGCCGACAGAGCTGCTGAGAGTACAGT 8619  
 Qy 2801 SerSerGlyAsnGlyAlaPro----- 2807  
 Db 8620 GAGGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8679  
 Qy 2808 ---GluGluArgLeuArgGluAsnGlyAspAlaLeuSerArgGluGlySerLeuGlyPro 2826  
 Db 8680 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8739





Db 1252 GAGGAGGGCGCGCAAGTACCAAGTCTCGTGGAGGCGCAAGCAAGCAAGGGCGCAATCCGGGC 13111

Qy 378 ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaPro 397

Db 1312 CCGCTCAATGCGCAGGCGCCACCGTGTACATCGAGGGGAGGAGCAAGAAAGAACTACATCC 13711

Qy 398 GluPheSerGluLysArgTyrValValGlnValArgGluAspValThrProGluYalaPro 417

Db 1372 CAGTTTCAGCGAGCAGAACTACTGTGTCCAGGTGCCCGAGAGCACTGGGGCTCAACAGCGCT 14311

Qy 418 ValLeuArgValThrAlaSerAspArgAspLysGlySerAspAlaValAlaHisTyrSer 437

Db 1432 GTGCTGCGAGTGCAGGCGCCAGCGGACCGAGGCGGAGAACCGCGGCACTTCACTACAGC 14911

Qy 438 IleMetSerGlyAspAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyYalaLeuAsp 457

Db 1492 ATCTCAAGCGGAGAACGTGGCCGGCCAGTTCTTACCTGCACTGCGTGAAGGGGAGATCCGTGAT 15511

Qy 458 ValValSerProLeuAspTyrGlnThrThrArgGluTyrThrLeuArgValAlaGln 477

Db 1552 GTGATCAACCCCTTGATTTGAGAGATGTCAGAAATATCGCTGAGCATTAAGGCCAG 16111

Qy 478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497

Db 1612 GATGGGGGCGGGCCCGCCCTCATCAATCTTCAGGGGGGTGTGTGTGTCAGGTCTGTGAT 16711

Qy 498 IleAsnAspAspAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer 517

Db 1672 GTCAACGACAAACGAGCTTATCTTTGTAGACAGCCCTTCCAGGCCACGGTCTGGAGAT 17311

Qy 518 ValProLeuGlyTyrTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537

Db 1732 GTGCCCTCGGGCTACCCCGTGGTGCATCATCAGGGGTGAGCAGCGGACTCTGGAGAGAAC 17911

Qy 538 AlaArgLeuGluTyrArgLeuAla-----GlyValGlyHis-- 549

Db 1792 GCCGGGCTGCACTATCGCTGGTGAGCAGCGCTCCACTTTCGGGGGGCGGACGGCT 18511

Qy 550 -----AspPheProPheThrIleAsnAsnGlyTyrGly 560

Db 1852 GGGCTTAAGATCTGCGCCCAACCCCTGACTTCCCTTCCAGATTCACAAACAGCTCGGT 19111

Qy 561 ThrPheSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580

Db 1912 TGGATCAACAGTGTGTGCTCGAGCTGGACCGCAGAGGTGGAGCACTACAGCTTGGGGTGT 19711

Qy 581 GluAlaAspAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600

Db 1972 GAGCGCGTGAACACAGCGCTCGCCCCCATGAGCTCTCCACAGGTGTCAATCAGGTG 20311

Qy 601 LeuAspValAsnAspAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620

Db 2032 CTGGACGTGAATGACAAAGCAACCGGTGTTCACGACCCCACTACAGACTCGTCTGAAT 20911

Qy 621 GluAspAlaAlaValGlyTyrSerValValThrValSerAlaValAspAspAlaHis 640

Db 2092 GAGAGTGGCGCGGTGGGAGCAGCGCTGTACACCTGCAAGGCCCGGACCGTGAAGCCCAAC 21511

Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660

Db 2152 AGTGGATTACTTACCACTGCACTGCAAGCGGCAACCCGGAACCGTTTGCACTGAGAGC 22111

Qy 661 GlnSerGlyGlyGlyLeuValSerIleuAlaLeuProLeuAspTyrTyrLeuGlnArgGln 680

Db 2212 CAGAGAGGGGGGGGCTTACTACCTCTGCGCTTCACTGCACTACAGACAGAGAGAGAG 22711

Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700

Db 2272 TACGTGCTGGGGGTACAGACATCCGACAGCGTGCACATCGGCGCAATGCTCTAATC 23311

Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720

Dd	2332	AAAGCTCACTAATGGCAACACCACAGGCGCTGTCTTTCAGAGCTCCATTACAGATGAGT	2391
Qy	721	ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740
Dd	2392	GTACGTATGAGCAGAGCGCTGTGGGCACTCTCAATGGCTCCCTCACTGTCCACAGAAAGAC	2451
Qy	741	ThrGlyGluAsnAlaArgIleThrThyPheMetGluAspSerIleProGlnPheArgIle	760
Dd	2452	ACAGAGAGAGATGGCCCCCATCACTTAAGTATTCAGAGACCCCGCGGAGTTCCGACTT	2511
Qy	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer	780
Dd	2512	GACCCCGCACAGTGGCACCATGTACACCATATGAGTGAAGCTGACTATATGAAACAGAGTGGCC	2571
Qy	781	TyrThrIleuAlaIleThrAlaArgAspAsnGlyIleProGlnIleuSerAspThrThrTyr	800
Dd	2572	TACAGCGTACCAATCAATAGGCCAGAGACAGGACTCCCGAGAAATACAGACCAACACC	2631
Qy	801	LeuGlnIleLeuValaAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820
Dd	2632	CTAAGAGATCCCTCAATCCCTGATAGCCAAATGACATGCCACCCGAGTTCCGTGGGAGATTCTAC	2691
Qy	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Dd	2692	CAGGGTTCCACTTTCATGAGATGCTCCACCTTCGACCGAGATCCCTCCAGGCTCTGCACAG	2751
Qy	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly	860
Dd	2752	GACCGGAGCTCAGTCCCATGGGCGCTCTGCTGACCTTCCAGGGTGGGAGACAGCGC	2811
Qy	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880
Dd	2812	GATGGGAGCTTCTTCAATCCAGGCCACCGCTCGGTGTGATTTGGCACCCGCGCTGTGAC	2871
Qy	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetPro---	899
Dd	2872	CGGAGAGATGTGGCGGTGACACCTTTGGCTCTGTGCTGTGATTCGGGGCATGCCACT	2931
Qy	900	ProAlaArgThrProMetGluValThrValThrValLeuAspValaAspAsnProPro	919
Dd	2932	CCCCTTACCGCGCTGGTAGAAATATCAGATGACCATCTTTCGACATTATATGACAAAGCCCCC	2991
Qy	920	ValPheGluGluAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAla	939
Dd	2992	ATGTGTGAAGAAGCAGACTGAGCGCTGTGTGTGAAGAGAAACAACCAAGTGGGGTCCGTG	3051
Qy	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln	959
Dd	3052	GTGGCAAAAGATCGTGTTAACGACCCCTGATGAAGAGCCCTAATGCCAGATCATATATCAG	3111
Qy	960	IleValGluGlyAsnIleProGluValaPheGlnLeuAspIlePheSerGlyGluLeuThr	979
Dd	3112	ATTGTGTGAAGGGGACATGCGCGCATTTCTTCAGCTGACCTGCTCCATCAAGGGGACCTTCGT	3171
Qy	980	AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr	999
Dd	3232	TGGCTCCGCTGGTGAAGCCAGGCCAGGTCACATCTTCTTCGGGACCAAGATATGACAC	3291
Qy	3172	GCCATGTGTGAGCTGTGACTTTGAGGCTCCGGCGGAGATATGTCTGTGTGTGACGACAGC	3231
Dd	3232	GGCCTGTGTGAGCTGTGACTTTGAGGCTCCGGCGGAGATATGTCTGTGTGTGACGACAGC	3291
Qy	1000	SerAlaProLeuValSerArgAlaThrValIleValaArgLeuLeuAspArgAsnAspAsn	1019
Dd	3292	TCGGCTCCGCTGGTGAAGCCAGGCCAGGTCACATCTTCTTCGGGACCAAGATATGACAC	3351
Qy	1020	ProProValLeuGlyAsnAspGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer	1039
Dd	3352	AGTTTCCCAACCGGCGTGAATCGGCTGATCCCGGCCATGACCCCGAGCGTGCAGACAGC	3411
Qy	1040	SerPheProGlyIleAlaIleGlyArgValProAlaIleAspProAspIleSerAspSer	1059
Dd	3412	CTCAACTACACCTTTCGACAGGGGACAGACCTGTGGCTGTGTGTGACCCCGCACAGC	3471

QY 1080 G1YGLIleuLYseuSerArgAlaIleuAspAsnAspArgProLeuGlnAlaIleMetSer 1099  
 DB 3472 GGGGAACTGAGCTCAGCCGCGACCTGGACAACAACCGCCGCTGAGGCGCTCAATGAG 3531  
 QY 1100 ValLeuValSerArgpLYValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 DB 3532 GTGTCTGTCTGATGATGATCCACAGCTCAACGCTTCTGACCCCTGGGTGACATC 3591  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGlnAspMetSerProGlu 1139  
 DB 3592 ATCAAGAGAGACATGCTGACCAACAGATCACTGCTCCGTGAGAAACATGCTCCAGAG 3651  
 QY 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrIleAlaThr 1159  
 DB 3652 AGGTTCTGTGCTCCGCTGCTGCGCCCTTCTGTGAGAGGGGTGCGCCGCTGTCTCAAC 3711  
 QY 1160 ProProAspHisValValValPheAsnValGlnAspThrThrAspAlaProGlyGlyHis 1179  
 DB 3712 ACCAAGAGACGATCTTCTGCTTCAACGTCCAGAAACGACACCGACGTC---AGCTTCAAC 3768  
 QY 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhe 1199  
 DB 3769 ATCTGAAGATGACCTTCTGCGCGCTGCTGCGCGCGCGCGCGCGC-----CAGTTC 3822  
 QY 1200 LeuProSerGluAspLeuGlnArgLeuYrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 DB 3823 TTCCTGTCGAGGAGCCTGCGAGAGAGAGATCTACATCGACGCTGCTGACCAACATC 3882  
 QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
 DB 3883 TCCAGCGACGCGCGCTGCTCCCTCGACGAACAATCTGCTCGCGACGCGCTGCGAGAAC 3942  
 QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
 DB 3943 TACATGAAGATGGTGGTCCGTTCTGCGATTCGACAGCTCCGCGCTTCTCTCAAGCTTCAC 4002  
 QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
 DB 4003 ACCGATCTCTTCGCGCCATCCACCCCAACAGCGCTGCGCTGCGCGCTGCGCGCGC 4062  
 QY 1280 PheThrGlyAspTyrCysGluThrGlyValAspLeuCysTyrSerArgProCysGlyPro 1299  
 DB 4063 TTCACCGCGGACTACTGCGACAGAGATCGACTCTGCTACTCCGACCCGTCGCGCGC 4122  
 QY 1300 HisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
 DB 4123 AACGCGCTGCTGCGAGCCGACGCGAGGCGGAGCTACCTGCGAGGTCTTCAAGGACTTCACT 4182  
 QY 1330 GlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsn 1339  
 DB 4183 GGAAGACACTGTGAGGTGATCCCGCTCAGGCGCGCTGCGCAAGGGGTGTGCAAGAAC 4242  
 QY 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheIleCysAspCysProSerGlyAsp 1359  
 DB 4243 GGGGGCACTGTGTAACCTGCTCACTGCGCGCTTCACTGCTGTGTCTCTGCGGAG 4302  
 QY 1360 PheGluIlyProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
 DB 4303 TATGAGAGCCCTTACTGTAGAGGTACACACAGAGGTTCGCCCCAGCTCTTCTGTCAAC 4362  
 QY 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGlu 1399  
 DB 4363 TTCGGGGGCTGAGACAGGCTTCCACTTCAACATCTCCCTCAGCTTGGCACTGAGAA 4422  
 QY 1400 ArgAspGlyLeuLeuLeuThrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeu 1419  
 DB 4423 AGGAACGGCTGTCTCTCAACAGCGCTTCAATGAGAACAGACTTCACTCCCTG 4482  
 QY 1420 GluValIleGlnGluValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
 DB 4483 GAGATCGTGAACGACAGGTGACCTCACTTCTGTGAGGAGAACACACAGACGCTG 4542

QY 1440 SerProPheValProGlyValSerAspGlyGlnThrPheIleThrValGlnLeuIlyTyr 1459  
 DB 4543 GCACGAAAGTTCCAGTGTGTAGTACGGGCGGTGACCTGTGTGAGGTGACATC 4602  
 QY 1460 TyrAsnIlyProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIlyVal 1479  
 DB 4603 TACAACAAGCCCAATATTGGCCACTGGGCGCTGCCCAATGGGCGCTCGGGGAAAAAGATG 4662  
 QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 DB 4663 GCGGTGTGACATGTGATGATGTGACAAACATGAGGTGTGCTTGGAAAGACATC 4722  
 QY 1500 GlyAsnIlySerCysAlaAlaGlnGlyThrGlnGlyGlySerIlySerLeuAspLeu 1519  
 DB 4723 GGGAACTACGCTGCGCTGCCAGGCACTGACCGGCTCCAAAGATCCTGTGATCTG 4782  
 QY 1520 ThrGlyProLeuLeuLeuGlyValProAspLeuProGluSerPheProValArgMet 1539  
 DB 4783 ACCGCGCTTACTCTCGGGGGGTGTCCCAACCTGCCAAGACTTCCAGTGCACAAAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
 DB 4843 CGGCAGTTCGTGGGCTGATGCGGAACCTGTCAAGTCGACGGCAAAATGTGACATGCCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaIlyIlyAsnValCysAsp 1579  
 DB 4903 GGATTCATCGCCACAAATGGACCCCGGAGAGCTGCTGCTGAGGAACCTTGTGCAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyIlyThrCysValAsnGlnTyrAspAlaPheSerCysGlu 1599  
 DB 4963 GGGAGGCGGTGAGAAATGAGAGGACCTGTGTCAACAGTGGAAATATGATCTGTGTAG 5022  
 QY 1600 CysProLeuGlyPheGlyIlyIlySerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTCCATCTCCATTTCCGGGGAGAGACGTGAGAACGATGCTCAACCCCACTCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaThrPheIleGlyLeuSerLeuProIleSerGlyProThrTyr 1639  
 DB 5083 AGCGGTGAGACGTGTCTGTCTGAGATGACCTGCAACATCACTCTGTGTGCTGCTGATC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 DB 5143 CTGGGCTCATGTTCCGAGCCCGGAGAGGACAGCGTTGTGATGAGGACCAAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGly 1679  
 DB 5203 GGGCCCAACGACTTTCGCTTCAATCTCTGAACACTCACTTCAAGTTGAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThr 1699  
 DB 5263 GGGCCCTCCATGTGAGATCCGTATGCTGTCCGGTTGCCGTGACCGAGGAGATGG 5322  
 QY 1700 HisIleAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 DB 5323 CACCACCTGCTGATCGAGCTGAGAAATGTTAAGAGACAGTGAATGAACACTGGTTC 5382  
 QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGluIlyAsnLeuGlyProArgLeuHis 1735  
 DB 5383 ACATGAACCTTGGACTATGGAGTGAACCAAGAACAGATATCGGGGAGATCTTCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyIlyIleProGlyProAlaGlyValAla 1755  
 DB 5443 GGGCTGACGATGAAGACGTGTGTGTGGAGGCCCTCTGAAGACAGTTCCTGTCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGluGlyValAsn 1775  
 DB 5503 CGTGAATTCGAGGCTCATAGCAGGAGTGAAGATGGGGGGAGCGCCACCAAGTGC 5562  
 QY 1776 SerLeuAspProSerHisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAsp 1795  
 DB 5563 ACCTGAACATGAACAAACGACTCAAGTCAAGGTGAAGACGCTGTGATGTGACGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815





QY 32 pGlnValGlyProCysArgSerLeuGly---SerArgGlyArgGlySer----- 47  
 Db 59 -----CCCTGCGCGGAGATGAGGCTGCGGAGCGCGCTGAGAGCGCGGTACCC 108  
 QY 48 -----SerGlyAlaCysAlaPrometGlyTrp-----Le 57  
 Db 109 GCGGAGACCGCGGCTTCGCGCTTCGCGCGCTGTAACCTTACGCGGTGAGCGCGCTTGC 168  
 QY 57 uCysProSerSerAlaSerAsnLeuTrpLeuTrpSer----- 70  
 Db 169 AGCGCCCGGAGCGCGGAGAGCTGAGAGCTGAGCGCGAGATGAGGCGGCTGAGAGAGCT 228  
 QY 71 -----ArgCysArgAspAla-GlyThrGlnLeu 80  
 Db 229 CGCGCGCTTCGCGCGCGGAGCGCGCTGCGCTGCTCAAGTCCGCTTGGT----- 278  
 QY 80 hrcGlyHisLeuValProHisAspGlyLeuArgValTrpCysProGlnSerGlnAlaH 100  
 Db 279 --GCGCCGAGTGCCTCC--GACGAGGCTGAGCGG-----CGCGCTGCGGCGCG 324  
 QY 100 IsrLeProLeuProProAlaProGln-GlyCysProTrpSerCysArgLeuGlyIle 119  
 Db 325 AC-----GACCTTCCCGGCTGCGAGCCCGTACCGCGCTTGCAGAAC 369  
 QY 120 GlyGlyHisLeuSerProGlnGlyLysLeuThrLeuPro----- 132  
 Db 370 GGTGCGCGGCTGCG-----GGGAGCGCTTCTGCTTCCGCTCCGCGGCTGCGGCGC 423  
 QY 133 GlnGlnHisProCysLeuLysAla----- 140  
 Db 424 GCGGACATTCGCGCGCTTCGAGCTCCGACACCTTACCGCGCTGCGCGCGCGC 483  
 QY 141 ProArgLeuArgCysGlnSer----- 147  
 Db 484 CCAGGCGCGCGCTGCTCCGCGCGCTCCCATCTGCTGCGCGCGGCGCTGCGCGCTG 543  
 QY 148 -----CysLeuLeuAlaGlnAlaProGly---LeuArgAlaGly----- 159  
 Db 544 CGTCTGCTGTGCGCTGCGCGCGCGCGCGCTGCGCGCTGCGGTGAGACTGCGCTGAG 603  
 QY 160 -----GluArg 161  
 Db 604 GCGGACCGCGGAGACCGCTTCGCGCTGCGCATCCCATCGCGCGCTTCGCGCGGAC 663  
 QY 162 SerProGlnGlnSerLeuGlyArgArgGlyArgAsnValAsnThrAlaPro----- 179  
 Db 664 TTGCGCGGAGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGG 723  
 QY 180 -----GlnPheGlnProProSerTrpGlnAlaThrValProGlnAsnGlnProAlaGly 197  
 Db 724 AGCTGTAAGTTCCTCGATGCGCCCACTACAGGTGGCGCTTGTGTGAGAACAGACCGGCGGCG 783  
 QY 198 ThrProValAlaSerLeuArgAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGln 217  
 Db 784 ACCCTCATCTCCAGTTCAGCGCGCACTACACATGAGGCGAGAGAGAGAGCGCGTGAAC 843  
 QY 218 TyrThrMetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProVal 237  
 Db 844 TATTATATGAGAGGCGCTGTTGACGAGCGCTCCGCGGCTACTTCCGATTCAGATTCCTGCGC 903  
 QY 238 ThrGlyAlaValThrThrAlaGlnGluLeuAspArgGlnThrLysSerThrHisValPhe 257  
 Db 904 AGCGGCGCGGTGAGCGAGACAGCGTACTGAGCCGAGAGACCAAGAGAGCGACGCTCTC 963  
 QY 258 ArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle 277  
 Db 964 AGGTGAAGCGCGTGAAGCTACAGTACGCGCGCGCTCGCGCACCACTACATCACTGTC 1023  
 QY 278 LeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnGlnLysLeuGlnSer 297  
 Db 1024 TTGTGTAAAGACACCAACAGACCAACCGCGCTTGTGACAGTCCGAGTACCGCGAGCGC 1083  
 QY 298 LeuArgGlnAsnLeuGlnValGlyTrpGlnValLeuThrValArgAlaThrAspGlyAsp 317

Db 1084 GTGCGGAGAGACCTTGAGAGTGGGCTACAGAGGTGTCACATCCGCGCGAGAGCGCGAC 1143  
 QY 318 AlaProProAlaAsnIleLeuTrpArgLeuLeuGlnGlySerGlyGlySerProSer 337  
 Db 1144 TGCCCATCAACGCCAATCTTGATTCGCGCGTGTGGGGGCGCTGCG----- 1191  
 QY 338 GlnValPheGlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArg 357  
 Db 1192 GACGTCTTCAGACTCAACGAGAGCTCTGCGCGGTGAGACACAGCGCGCGGTGTGACCGG 1251  
 QY 358 GlnGlnValGlnSerTrpGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGly 377  
 Db 1252 GAGGAGCGCGCGAGTACCACTCTGCTGAGGCGCAACGACAGCGCGCGCATCCGGGC 1311  
 QY 378 ProArgSerThrThrAlaAlaValPheLeuSerValGlnAspAspAspAsnAlaPro 397  
 Db 1312 CGCTCATGTGCCACCGGCCACCGTGTATCTCAGGTGAGAGAGAGAGACCACTACCTCC 1371  
 QY 398 GlnPheSerGlnLysArgTrpValValGlnValArgGluAspValThrProGlyAlaPro 417  
 Db 1372 CAGTTCAGCTGAGACAGAACCTACGTGGTCCAGGTGCCGAGAGAGTGGGCTCAACCGGT 1431  
 QY 418 ValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValAlaHisTrpSer 437  
 Db 1432 GTGCTGCAGATGCAAGCCACGACCGGAGCCAGGCGCAAGCGCGCATTCATCAACAGC 1491  
 QY 438 IleMetSerGlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyAlaLeuAsp 457  
 Db 1492 ATCTTCAGCGGAGAGAGTGGCGCGCGCATCTTCACTGCTGCTGAGCGGAGTCTGGAT 1551  
 QY 458 ValValSerProLeuAspTrpGlnThrThrLysGlnTrpThrLeuArgValArgAlaGln 477  
 Db 1552 GTGATCAACCCCTTGAGATTCGAGAGATGTCAGAAATCTCGTGAAGATTAAGCCGAG 1611  
 QY 478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
 Db 1612 GATGGGAGCGCGCGCGCTCATCAATCTTCAGGAGGTGTGTGTGTGTGTGTGTGTGTGT 1671  
 QY 498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
 Db 1672 GTCAACAGCAACAGACCTTCTTGTGAGACGCCCTTCAGAGCCACGCTGTGAGAT 1731  
 QY 518 ValProLeuGlyTrpLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
 Db 1732 GTGCGCGTGGCTACCCCGGTGTGACATTCAGGCGGTGAGACCGCGCATTCGTGAGAGAAC 1791  
 QY 538 AlaArgLeuGlnLysTrpArgLeuAla-----GlyValGlyHis--- 549  
 Db 1792 GCGCGGTGCACTATCGCTGTGTGAGACAGGCTTCACCTTCTGCGGCGCGCGCGCT 1851  
 QY 550 -----AspPheProPheThrIleAsnAsnGlyTrpGly 560  
 Db 1852 GGGCTTAAGATTCCTGCCCCCACCCCTGACTTCCCTTCAGATTCACAAACAGCTCGGCT 1911  
 QY 561 TrpIleSerValAlaAlaGlnLeuAspArgGlnGluValAspPheTrpSerPheGlyVal 580  
 Db 1912 TGATATCAAGTGTGTGCGGACTGAGCCGCGAGAGGTGAGGACCTACAGCTTCGCGGTG 1971  
 QY 581 GlnIleArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1972 GAGGCGGTGAGACAGGCTGCGCCCCCATGAGCTCTTCACAGGTGTTCATCAACGCTG 2031  
 QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlnLysTrpValArgLeuAsn 620  
 Db 2032 CTGACAGGTGAATGACAAAGACCCGCTGTTCACGAGCCACACTTACAGCTTCGTGTGAAT 2091  
 QY 621 GlnAspAlaAlaValGlyThrSerValValThrValSerAlaAlaAspArgAspAlaHis 640  
 Db 2092 GAGGATGCGCGGTGTGGAGAGCGGTGTGACCTCTGAGCGCGCGCGAGCCGTGACCCCAAC 2151  
 QY 641 SerValIleThrTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660



Db 2152 AGTGTGATTACTTACCAAGCTCAAGGCGCAACCCGGAACCGCTTTCGACTCAAGCAGC 2211  
QY GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleuGln 680  
Db 2212 CAGAGAGGGGGGGCTCATCACCTGGCGCTACCTTGGAATCAAGAGAGGAGCAGC 2271  
QY TyrValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
Db 2272 TACGGCTGCGGAGTACAGCATCCAGCAGCGCTCCGACCTCCGATGCTCCCAATC 2331  
QY AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
Db 2332 AACGTCACTGATGCCAACAACCCAGGCTGCTTTCAGAGCTCCCATTCACAGAGT 2391  
QY ValAsnGluAspArgProAlaGlyThrThrValIleuIleSerAlaThrAspGluAsp 740  
Db 2392 GTCAAGTACAGACAGGCTGTGGGCACTCCATTTGCTACCTCAGTCCAAAGATAGAGC 2451  
QY ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACAGAGAGAAATGCGCGCATCACTACGTATTCAGAGACCCCGTCCGACGTTCCGCAAT 2511  
QY AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
Db 2512 GACCCCGACAGTGGCACTGACATGACATGAGCTGAGCTGATGAGAACAGAGTCCGC 2571  
QY TyrThrIleuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db 2572 TACACGCTGACCATCATGCGCCAGAGCAACGCGCATCCGCAAAATCAGACACCAACC 2631  
QY LeuGluIleLeuValAsnAspValAsnAspAlaProGlnPheLeuArgAspSerTyr 820  
Db 2632 CTAGAGATCTCATCTCTGATGCCAATGACATGACCCCAAGTCTCTGGGATTTCTAC 2691  
QY GlnGlySerValTyrGlyuAspValProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGTTCCATCTTGGAGATGCTCACTTCGACACAGATCTCTCAGGTCTCTCCACG 2751  
QY AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860  
Db 2752 GACCCGGAGCTCAGATCCCAATGGCGTCTGTACACTTCAGAGGTGGGAGCAACGCGC 2811  
QY AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgLeuAsp 880  
Db 2812 GATGGGAGCTTCTACATCAGACCCAGTCCGGTGTGATTCGCAACCAAGCGCGCTGAGC 2871  
QY ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspIleGlyPhePro--- 899  
Db 2872 CGGAGGAATGGCGGTGACAACTTTGGGCTCTGGCTGTGGATCGGGGCAATCCCACT 2931  
QY ProAlaArgThrProMetGluValThrValThrValIleuAspValAsnAspAspPro 919  
Db 2932 CCCCTTACCGCTCGGTAGAAATCCAGGTGACCATCTTGACATTAATGACAAATCCCCC 2991  
QY ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTTTGAGAGAGCAACTGAGCTGTTTGTGAGAGAACACCAAGTGGGGTCCGGTG 3051  
QY ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGGAAAGATTCCGCTTACAGACCTTGATGAGGCCCTTAATGCCAATTCATGATCAG 3111  
QY IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyValuLeuThr 979  
Db 3112 ATTTGGAGAGGGGAGCATCGCATTTCTCCAGCTGACCTGCAACGGGGGACCTGGCT 3171  
QY AlaLeuValAspLeuAspTyrGluAspArgProGluTyrThrValIleuValIleGlnAlaThr 999  
Db 3172 GCATAGTGGAGCTGACTTGAAGTCCGGCGGAGTATGTGCTGTGTGAGGCGCACG 3231  
QY SerAlaProLeuValSerArgAlaThrValHisValAlaGlyLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGCTGTAGGCGAGCCAGGTGACATCTTCTGTGTGACCAAGATACAAAC 3291

QY ProProValIleuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSer 1020  
Db 3292 CGCGCTGTGCTGCCGCACTTCAGATCTCTTCAACAACATATATGACCAACAAGTCCAAAC 3351  
QY SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGGCGTATGCGGTGATCCCGGCGCATGACCCGAGGTGTCAAGAC 3411  
QY LeuThrTyrSerPheGluArgGlyAsnGluLeuSerIleuValIleuAsnAlaSerThr 1079  
Db 3412 CTCAACTACCTTGTGTGAGGAGGCAACGAGCTGCCCTGTGTGCTGAGTGAACCCGCAAG 3471  
QY GlyGluLeuValSerAspArgAlaLeuAspAsnAsnAspProLeuGluAlaIleMetSer 1099  
Db 3472 GCGCAACTGACGCTCAGCGCGGACCTGGAACAACAACCGCGCGTGGAGCGCTCATGGAG 3531  
QY ValLeuValSer---AspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThr 1118  
Db 3532 GTGTCTGTGTGAGATGGATGCATCAGCGCTCAGCGCTTCTGACACCTCGTGTCAAC 3591  
QY IleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerPro 1138  
Db 3592 ATCATCAGAGACATGCTGACCAACAGCATCTGCTCCGCTGGAGAAACATGTCAG 3651  
QY GluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAla 1158  
Db 3652 GAGAAGTTCCGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3711  
QY ThrProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGly 1178  
Db 3712 ACCACCAAGAGCAGAGTCTTCTGCTTCAACGTCCAGAACAGACACCGAGCTC---AGCTCC 3768  
QY HisIleLeuAsnValSerLeuSerValGlyGlnProProGlyGlyGlyProPro 1198  
Db 3768 AACATCTGAACTGACCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3822  
QY PheLeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAla 1218  
Db 3823 TTCTTCCCGTGGAGGAGCTGCAAGAGAGATACCTACCTGATCCGAGCGCTGACCAAC 3882  
QY IleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGly 1238  
Db 3883 ATCTTCACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3942  
QY AsnTyrMetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleAlaSer 1258  
Db 3943 AACTCATGAATGCGTGTCCGTTCTGCGATTCAGACGCTCCGCGCTTCTCAAGCTCC 4002  
QY SerSerValIleuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProPro 1278  
Db 4003 ACCACCGTGTCTTCCGCGCATCCACCCATCAACAGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 4062  
QY GlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGly 1298  
Db 4063 GCGTTCAACCGCGCACTCTGAGACAGAGATCACTCTGCTACTCCGACCCGTTGGGCGC 4122  
QY ProHisGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyr 1318  
Db 4123 GCCAAGCGCGCTGCGGAGCGGAGCGGAGCGGAGCTCACTCCGAGTGTTCGAGAGATTC 4182  
QY ThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysAla 1338  
Db 4183 ACTGAGAGAGCACTGTAGGTGAGTCCCGCTCAGGCGCTGTGTGCAACGGGGTGTGCAAG 4242  
QY AsnGlyValThrCysValAsnLeuValGlyGlyPheCysAspCysProSerGly 1358  
Db 4243 AACGGGGCACTGCGTGAACCTGCTCATCGCGCGCTTCACTGCTGCTGCTGCTGCTGCTGCT 4302  
QY AspPheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIle 1378  
Db 4303 GAGTATGAGAGGCGCTTACTGTGAGTGAACAACGAGAGCTTCCGCGCCCACTCTTGTGTC 4362

QY 1379 ThPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLys 1398  
 DB 4363 ACCCTCCGGGGCCCTGAGACAGCGCTTCACCTTCACCTCCCTCAGGTGGCAGCTCAG 4422  
 QY 1399 GlnArgAspGlyLeuLeuLeuLysArgHisGlyPheAsnGlySerPheValAla 1418  
 DB 4423 GAAAGGAACGGCTGCTCTCTACACGGCCGGCTTCATGAAAGACAGACTTCATCGCC 4482  
 QY 1419 LeuGluValIleGlnGlnValGlnLeuThrPheSer-----AlaGlyGluSerThr 1436  
 DB 4483 CTGGAGATCCGAGACGAGGTGACAGCTCAGCTTCCTGAGGTGAGGAGAGACACA 4542  
 QY 1437 ThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrPheIleValGln 1456  
 DB 4543 ACGACCGTGGCAGCGAAGGTCCAGGTGTGTGAGTACGCGCGGTGGCACTGTGTGAG 4602  
 QY 1457 LeuLysPheTyrAsnLys-----ProLeu 1464  
 DB 4603 GTGCGAGTACTACAAAGGTAAATGGGCCCCACCACTTCCCTGGCCCCCAGCCCAAT 4662  
 QY 1465 LeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLysValAlaValAlaThrVal 1484  
 DB 4663 ATTGGCCACCTGGGCTGGCCCATGGCCGTCCGGGAAAAGATGGCGGTGTGACAGTG 4722  
 QY 1485 ArgGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGlyAsnTyrSerCys 1504  
 DB 4723 GATGATTTGTACACAACTGGCTGTGCTGTGGAAAGACATCGGGAACCTACAGCTGC 4782  
 QY 1505 AlaAlaGlnGlyThrGlnGlyGlySerLysSerLeuAspLeuThrGlyProLeuLeu 1524  
 DB 4783 GCTGCCAGAGGACTCAGACCGGCTCCAAAGATCCCTGAATCTGACCGGCTCTACTC 4842  
 QY 1525 LeuGlyGlyValProAspLeuProGluSerPheProValArgMetArgGlnPheValGly 1544  
 DB 4843 CTGGGGGGGTGCCCACTGCCAGAGACTTCCCACTGCAACAACGGAGTTCGTGGGC 4902  
 QY 1545 CysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAspPheIleAlaAsn 1564  
 DB 4903 TGCATGGGAACCTGTCACTGTGACGGCAAAATGTGACATGGCGGATTCATCGCCAAC 4962  
 QY 1565 AsnGlyThrValProGlyCysProAlaLysLeuValAsnValCysAspSerAsnThrCysHis 1584  
 DB 4963 AATGGACCCCGGAAGGCTGGCTGCCTCGAGAACTTCTGCATGGGAGCGGTGTGACG 5022  
 QY 1585 AsnGlyGlyThrCysValAsnGlnIleAsnIleProValAspSerCysGluCysProLeuGlyPhe 1604  
 DB 5023 AATGGAGGACCTGTGTCAACAGGTGAAATGTATCTGTGTAGTGTCCACTCCGATTC 5082  
 QY 1605 GlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeuGlySerSerLeu 1624  
 DB 5083 GCGCGGAAGAACTGTAGCAAGGCATGCTCAACCCCAAGCTTCAGCGGTGAGAGCGTTC 5142  
 QY 1625 ValAlaThrPheIleGlyLeuSerLeuProIleSerGlnProThrTyrLeuSerLeuMetPhe 1644  
 DB 5143 GTGTCTCGAGTACCTGAAACATCATCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 5202  
 QY 1645 ArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIle 1664  
 DB 5203 CGGACCGGGAAGGAGGACAGCGTTCATGAGGAGCCACAGGTGTGGGCCACAGGCTTC 5262  
 QY 1665 ThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGluGlyThrGlyLeuGlnAla 1684  
 DB 5263 CGCTCCAGATCTGAACAACTACCTCAGTTGAGGTGTCCACGGCCCTCCGATGTG 5322  
 QY 1685 SerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHisAlaGlnLeu 1704  
 DB 5323 GAGTCCGTAATGCTGTCCGGTGTGGGTGTACCGAGGAGTGTGCAACCTGTGATC 5382  
 QY 1705 AlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle-----LeuSerPheAsp 1720  
 DB 5383 GAGCTGAAGATGTTAAAGAGACAGTGAAGTGAAGACCTGTGCAACATGACCTTGTGAC 5442  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleLeuSer 1740

DB 5443 TATGGATGACCAAGAAACAGGACATATCGGGGGCATCTCCCGGGCTGACGGTAAG 5502  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 DB 5503 AGGTGTGTGTGAGAGCCCTCTGAAACAAAGTCTCCGTGGCCGTGTGATTCGAGGC 5562  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 DB 5563 TGCATCAGAGGATGAGATGGGGGGGAGCCCAACCAACTGCGCCACTGAACATGAAC 5622  
 QY 1781 HisGlyGluSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB 5623 AAGCACTCAAGTCAAGGTGAAGACCGCTGTGATGTAGACGACCCCTGTACTTGACG 5682  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyrSerCysSerCysAsp 1820  
 DB 5683 CCTGTCCCCCAATAGCCGCTGACCAACGCGCTGGAGAGACTACAGCTGCTGTGAC 5742  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840  
 DB 5743 AAAGGTACTCTGAATAAATGTGTGATGCTGTCACTGAACCTTGCGAGAACATG 5802  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 DB 5803 GGGGCTGTGTGCTGCCCGGCTCCCGACAGGCTGATGTGTGCGAGTGTGGCCAGT 5862  
 QY 1861 TyrLeuGlyProTyrCysArgLysThrArgIleAspGlnProCysProArgGlyTyrThrGly 1880  
 DB 5863 CACTAGGGGCTGTACTGTGAAACAACTGCACCTTCCGTGCCAGAGGCTGTGTGGGG 5922  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn 1900  
 DB 5923 AACCCGCTGTGAGACCTGTGCACTGTGTGCGCAAGAAAGCTTGTGATCCCACTGTAAAT 5982  
 QY 1901 LysThrSerGlyGlnCysHisGlyLysGluAsnHisTyrArgProProIleSerProThr 1920  
 DB 5983 AAGACCAAGCGGCAGGTGCMAATGCAAGAAATTACTCAAGCTCTTACGCCAGACACC 6042  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940  
 DB 6043 TGTCTCCCTGTGAGTCTGCCCATGGCTCCACAGCGGACGACTTGCAATGGCCACC 6102  
 QY 1941 GlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 DB 6103 GGGCACTGTGCTGTGCAAGCCCGGTGCATCGCGCCAGTGTGAACCGCTGTGCAACCCG 6162  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysAspProArgAlaIle 1980  
 DB 6163 TTTCGGAAGTCAACGCTCGGCTGTGAAGTACTACAAATGGCTGTCCAAAGCATTT 6222  
 QY 1981 GlnAlaGlyIleThrProArgThrArgPheGlyLeuProAlaAlaAlaAspCysPro 2000  
 DB 6223 GAGCGCGGATCTGTGTGCGCAAGACCAAGTTCCGAGACCGGCTGTGCGCATGCCCT 6282  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArgGlyTyrLeuProPro 2020  
 DB 6283 AAGGAGTCCGTTGGAAATCCGCTCCGACACTCAGCGGGAAGAGGCTGTGCTGCCCA 6342  
 QY 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu 2040  
 DB 6343 GAGCTTTTAATCTGACCAACTCTCTCTGTGAGACTCAAGGCCATGAATGAAGCTGTG 6402  
 QY 2041 GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
 DB 6403 AGCGGATAGAGACGAGGTGGAGCGGCGCCAGGCGCTGCAGCTGTGTGAGGGCGTGGCC 6462  
 QY 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
 DB 6463 AGTGCTAACACACACCGGACGCTCTTGTGCAATGAAGTGTGCGACGGCTTACAGCTG 6522  
 QY 2081 AlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGlyLysSerAlaThrGln 2100

Db 6523 CTGGGCGACGCTCTTCAGCAGAGAGCTGGCAGAGGCGCTTCGACCTGGCAGCGCAG 6582  
Qy 2101 AspValHisPheThrGluLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6583 GAGCGGACATTTACAGAGAGCGTCATCCGCGGAGCGCGCTCCGCGGCCAGCCAGC 6642  
Qy 2121 LysArgHisPheProGluLeuLeuIleGlnGlnThrGluGlyGlyThrAlaPheLeuLeuGlnHis 2140  
Db 6643 AGGAGCGCGTGGAGAGATCCAGCGAGCGAGCGAGCGCGCAGCGCAGCTGCTCCGCGC 6702  
Qy 2141 TyrGluAlaTyrAlaSerAlaLeuAlaGluAsnMetArgHisPheThrTyrLeuLeuSerProPhe 2160  
Db 6703 CTGAGGCGGCTACTTCACAGACGTCGACGAGCGAGCGCGGAGCGAGCGTCCGCGCCCTTC 6762  
Qy 2161 ThrTyrLeuValThrProAsnIleValIleSerValValArgLeuAspLysGlyAsnPheAla 2180  
Db 6763 GTCATCGTCACCGCCCAACATGTTCTGTGTGACATCTTTGACAAAGTTCAACTTTACG 6822  
Qy 2181 GlyAlaAlaLeuLeuProArgTyrGluAlaLeuArgGlyGlyGlnProProAspLeuGluThr 2200  
Db 6823 GAGGCGAGGGTCCCGGATTCGACACCATCATGAAAGTTCCCGAGGAGCTGGAGTCC 6882  
Qy 2201 ThrValIleLeuProGluSerValPheArg-----GluThrProProValVal 2216  
Db 6883 TCCGTCCTCTCCACGCGCACTTCTCAGACCACTGAGAGAAAGAGAGCGCCCTGCTG 6942  
Qy 2217 ArgProAlaGly-----ProGlyGluAlaGlnGluPro 2227  
Db 6943 AGGCGCGCTGGCGGAGACCAACCGCGAGACAGCGCGCGCTGGCGCACCGAGAG 7002  
Qy 2228 GluGluLeuAlaArgArgGlnAlaGArgHisPheGluLeuSerGlnGlyAlaValAla 2247  
Db 7003 GAGGCGCGCATCGACGAGCGGAGCGCACCCCTGATGACGCTGGCGCAGTTCGCGCT 7062  
Qy 2248 SerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProAspLys 2267  
Db 7063 CTGCTGATCATTTTACCGCACCCCTGGGAGACTCTCTCCGAGCGCTACAGACCCCGACCG 7122  
Qy 2268 ArgSerLeuArgValProLysArgProIleIleAsnThrProValValSerIleSerVal 2287  
Db 7123 CGCAGCCTCGGTTGCTCAACCGCCCATCATTAATACCGCATGTGAGCAACCGCGT 7182  
Qy 2288 HisAspAspGluGluLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArg 2307  
Db 7183 TACAGGAGAGGCGCTCCGCTCCGAGACCCCTGAGAGGCGCTCTGCGTGAAGTTGCGC 7242  
Qy 2308 LeuLeuGluThrGluGluArgThrLysProIleCysValPheThrAsnHisSerIleLeu 2327  
Db 7243 CTGCTGAGGTGAGAGAGACCAAGCCCTGTGCGTGTCTGGAAACCTCCCTGCGC 7302  
Qy 2328 ValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAsnGluSer 2347  
Db 7303 GTTGGTGGAGCGGAGGAGGTGTGCTGCCGCGGCTGGAGCTCTGTCAGAGAACCGGACA 7362  
Qy 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArg 2367  
Db 7363 CATGTGCTGTCAGAGTCAGACCAAGCCAGCTTTCGCTGCTCATGATATCTCCAG 7422  
Qy 2368 ArgGluAsnGlyGlyIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThr 2387  
Db 7423 CCGTGAAGACGGGAGGCTCTGCTCTGAAGATGTCACTATGCCGTGCTGTGTCA 7482  
Qy 2388 LeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArgGlyLeuLeuArgSerAsn 2407  
Db 7483 CTGGCGAGCGCTGTGGCTTCGCTCTGAGAGCTGTGCGGCAATGCTGGCGCTCAAC 7542  
Qy 2408 GlnHisGlyTyrLysArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu 2427  
Db 7543 CTGCAAGCAATTCACAGACCTGCGCGTGGCGCTTCCTCTCATGCTGTGTGCG 7602  
Qy 2428 LeuGlyTyrAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis 2447  
Db 7603 ATTGGATGATACCAAGACGAAACCGCTTCTGTGACAGTGTGCCATCTCTCTCAC 7662

Qy 2448 PheLeuTyrLeuCysThrPheSerTyrAlaLeuLeuGluAlaLeuHisPheTyrArgAla 2467  
Db 7663 TACATTTACATAGACACTTTCCTGAGCCCTGTGTGAGAGCTGTGATCCGCAAG 7722  
Qy 2468 LeuThrGluValArgAspValAsnThrGlyProMetLysPheTyrTyrMetLeuGlyTyr 2487  
Db 7723 CTGACCGAGGTGCGACATCAACAGCGGGCCCATCGGTTCTTCACTACGTCGTGGCTGG 7782  
Qy 2488 GlyValProAlaPheLeThrGlyLeuAlaValGlyLeuAspProGluIleTyrGlyAsn 2507  
Db 7783 GGATCCCGGCATTTCTACAGAGACTGGCGGTGCTGAGCTGAGCCCGCAGGCTACGGAAC 7842  
Qy 2508 ProAspPheCysTyrLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProVal 2527  
Db 7843 CCGACTTGTGCTGGCTGTGCTTCAAGACACCTTATTTGAGCTTTCGGGGCCCAAC 7902  
Qy 2528 AlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla 2547  
Db 7903 GAGCTGTATATATCATCAACAGTCACTTCTGTCTTATCTGAAAGGTTCTCTGCCAA 7962  
Qy 2548 AlaGlnArgGlnGlyPheGluLysGlyProValSerGlyLeuGlnProSerPheAla 2567  
Db 7963 ACAAAGCACATTATTATGGAAGAAAGATGCTCTCTGTGAGACCGCATTCCTC 8022  
Qy 2568 ValLeuLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeuLeuSerValAsnSerAspThr 2587  
Db 8023 CTGCTGCTGCTATACAGCGCCACCTGCTGCTGGGCTGCTGCTGTGAACCGCATGCA 8082  
Qy 2588 LeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
Db 8083 CTGAGCTTCACTACTCTCTCCCATCTTCAAGCGGCTTCAAGAGGCCCTTCTCTCCCTT 8142  
Qy 2608 SerTyrValValLeuSerLysGluValArgValAlaLeuLys---LeuAlaCysSerArg 2626  
Db 8143 TTCCACTGCTGCTCAACAGAGGCTCGAAGAGCACTGAAGGCGGTGCTCGCGGAGAG 8202  
Qy 2627 LysProSerProAspProAlaLeuThrThrLysSerThr---LeuThrSerSerTyrAsn 2645  
Db 8203 AAGCTGACCTGAGAGATCCGCGCACACAGGCGCACCTGTGACCGGCTCCCTCAAC 8262  
Qy 2646 CysProSerProTyrAlaAspGly-----ArgLeuTyrGlnProTyrGlyAspSerAla 2663  
Db 8263 TGCACACCACTTGCTGAGCGGCGCTGACATGCTGCACAGACTTGGCGAGTCCAC 8322  
Qy 2664 GlySerLeuHisSerThrSerArgSer----- 2672  
Db 8323 GCCTGCTGAGACGATGCTCAGGATGAGAGATCCAGAACTGCGCGTCTCTGGG 8382  
Qy 2673 -----GlyLysSerGlnProSerTyrIleProPheLeuLeuArgGlu 2686  
Db 8383 CTGCTGAGGGGAGCAGCAGAGAGCGAGCGGTCCTCAATGCCAGAGCGCAAGAT 8442  
Qy 2687 GluSerAlaLeuAsnProGlyGlnGlyProProGlyLysAspProGlySerLeuPhe 2706  
Db 8443 -----CCCCCTGGC----- 8451  
Qy 2707 LeuGluGluGlnAspGlnGlnHisAspProAspThrAspSerAspSerLeu 2726  
Db 8452 -----CAAGATTCAGCTCAGATGAGAGGTGCTCCCTG 8484  
Qy 2727 GluAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGluGluGlu 2746  
Db 8485 ---GATGACAGAGAGAGCTTTAGCGCTCTCAACACTGTCAAGACAGAGAGCATGGG 8541  
Qy 2747 GluGluGluGluGluGluAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGly 2766  
Db 8542 GTGGAGCTGAGGAAA-----TGGGACCGCGCGCAGGGGC 8577  
Qy 2767 ProGlyAlaGluArgLeuProLeuHisSerThrProLys---AspGlyGlyProGlyPro 2785  
Db 8578 -----GCCGTCAAGACCCCAAGGAGACGCTGTGCGAACAC 8619

[illegible]

```

RESULT 12
US-10-131-409-13
; Sequence 13, Application US/10131409
; Publication No. US20030199465A1
; GENERAL INFORMATION:
; APPLICANT: Malyanar et al.
; TITLE OF INVENTION: No. US20030199465A1 Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-67SCIPCON1
; CURRENT APPLICATION NUMBER: US/10/131,409
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 09/898,954
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/182,733
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/182,724
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/183,896
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 60/184,497
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/224,157
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/184,482
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,744
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/197,083
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/233,405
; PRIOR FILING DATE: 2000-09-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; TYPE: DNA
; LENGTH: 9087
; ORGANISM: Homo sapiens
US-10-131-409-13

Alignment Scores:      0
Pred. No.:
Length: 9087

```

Score:	8736.50	Matches:	1118
Percent Similarity:	69.08%	Conservative:	422
Best Local Similarity:	55.46%	Mismatches:	650
Query Match:	56.20%	Indels:	268
DB:	13	Gaps:	45

[illegible]

Db 904 AGCGGCGCGGTAGACAGCGGTACTGACCGCGAGACCAAGACGACGACCTCTC 963  
Oy ArgValThrAlaGlnAspHisGlyMetProArgSerAlaLeuAlaThrLeuThrIle 277  
Db 964 AGGGTGAAGCGGTGACTACAGTACCGCGCGCTCGGACCACTACCTACCTGTC 1023  
Oy LeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnGlnGlnGlnGln 297  
Db 1024 TTGGTCAAGACACCAACGACCAAGCGCGGTCTTCAAGCACTGGAATACCGCAAGCGC 1083  
Oy LeuArgGlnAsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAsp 317  
Db 1084 GTGCGGAGAACCTGAGGTGAGCTACGAGGTGACCACTCGCGGACCAAGCGCAC 1143  
Oy AlaProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSer 337  
Db 1144 TCGCCATCAACGCACTTGGCTTACCGCGGTGGGGCGCGCTG----- 1191  
Oy GlnValPheGlnValAspProArgSerGlyValIleArgThrArgGlyProValAspArg 357  
Db 1192 GAGGAGCGCGCGAGTACAGCTCTGCTGAGGCGCAACGACGAGCGCGCATCCGGC 1251  
Oy ProArgSerThrThrAlaAlaValPheLeuSerValGlnAspAspAsnAspAlaPro 397  
Db 1252 GAGGAGCGCGCGAGTACAGCTCTGCTGAGGCGCAACGACGAGCGCGCATCCGGC 1311  
Oy 378 ProArgSerThrThrAlaAlaValPheLeuSerValGlnAspAspAsnAspAlaPro 397  
Db 1312 CGGCTCAATGCGACCGCCACCTGTACATCGAGGTGAGAGAGAGAGAGAGAGAGAGAG 1371  
Oy GlnPheSerGlnValArgTyrValValGlnValArgGlnAspValThrProGlyAlaPro 417  
Db 1372 CAGTTCAACGAGCAGCACTACGCTGCTGAGGCGCGGAGGAGGAGGAGGAGGAGGAGG 1431  
Oy ValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValValHisGlyTyrSer 437  
Db 1432 GTGCTGCGAGGTGACGAGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAC 1491  
Oy IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaLeuAsp 457  
Db 1492 ATCTCAACGCGGAGAGTGTGCGCGCGGAGTCTACCTGCACTCGGAGCGGAGTCTGAT 1551  
Oy ValValSerProLeuAspTyrGlnThrTyrIleGlnGlyTyrThrLeuArgValArgAlaGln 477  
Db 1552 GTGATCAACCGCTTGTGATTCAGAGATGTCCAGAAATCTCGCTGAGCATTAAGGCCAG 1611  
Oy AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
Db 1612 GATGCGGCGCGCGCGCGCTCATCAATCTTCAAGGAGGTGTGTGTGTGTGTGTGTGTGT 1671  
Oy IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
Db 1672 GTCAACGACCAACGAGCTTCTTCTGTGAGAGCGCGCTTCAAGGACCGGTCTGTGAGAT 1731  
Oy ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
Db 1732 GTGCGCGCGGTGAGCGCGGTGTGACATTCAGCGGTGAGCGCGGATCTGTGAGAGAAC 1791  
Oy AlaArgLeuGlnGlyTyrArgLeuAla-----GlyValGlyHis--- 549  
Db 1792 GCGCGGCTGCACTATCGCTGTGTGAGACAGCGCTCACCTTTCTGTGGGCGCGCGCGCT 1851  
Oy 550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1852 GGGCTTAAGAACTCTGCGCGCGCGCGCTGATCTTCCCTTCAAGATCAACAAGCTCCGCT 1911  
Oy TrpIleSerValAlaAlaGlnLeuAspArgGlnValAspPheTyrSerPheGlyVal 580  
Db 1912 TGGATCAACAGTGTGTGCGAGCTGTGACCGCGGAGGTGTGAGCACTACAGCTTCCGGGTG 1971  
Oy GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerValSerValThrVal 600

Db 1972 GAGGCGGTGACCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2031  
Oy LeuAspValIleAsnAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsn 620  
Db 2032 CTGAGCTGTAATGACCAACGACCGCGGTCTTACGAGGCCACTTACAGACTTGTGTGAT 2091  
Oy GlnAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis 640  
Db 2092 GAGATGCGCGCGGTGGGAGAGCGGTGTGACCTTGAAGCGCGCGCGCGCGCGCGCGCG 2151  
Oy SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer 660  
Db 2152 AGTGTGATTCCTTACAGCTTCAAGCGCGCAACCGCGAACCGCTTGTGACCTAGCGAGC 2211  
Oy GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrValLeuGlnAspGln 680  
Db 2212 CAGAGAGGCGCGCGCTCATCACTTGGCGCTTCTGAGCTTCAAGAGAGAGAGAGAGAG 2271  
Oy TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
Db 2272 TACGTGCTGGCGGTGACAGCATCCGACGCGACAGGTGTGACACTGGCGCATGTCTATC 2331  
Oy AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
Db 2332 AACGTCACTGATGCGCAACCCACAGCGCTGTCTTTCAGAGCTCCATTCACAGTGAAT 2391  
Oy ValAsnGlnAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlnAsp 740  
Db 2392 GTCACTGAGAGAGAGCGCTGTGGGAGCATCTTCACTTCACTTCACTTCACTTCACTTCA 2451  
Oy ThrGlyGlnAsnAlaArgIleThrTyrPheMetGlnAspSerIleProGlnPheArgIle 760  
Db 2452 ACAAGAGAGATGCGCGCGCATCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 2511  
Oy AspAlaAspThrGlyValAlaValThrGlnAlaGlnLeuAspTyrGlnAspGlnValSer 780  
Db 2512 GACCCGACAGTGTGACCATGATGACCATGATGATGATGATGATGATGATGATGATGATG 2571  
Oy TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db 2572 TACAGCTGACCATCAAGCG 2631  
Oy LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSerTyr 820  
Db 2632 CTACAGATCTCTCATCTCGATGCGCAATGACATGACATGACATGACATGACATGACATG 2691  
Oy GlnGlySerValTyrGlnAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGCTTCATCTTGTGAGATGCTCAACCTTCAAGCAAGCATCTTCAAGCTTGTGACAG 2751  
Oy AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
Db 2752 GACCGGAGCTCAGGTCCCATGCGCGCTGTCTGTGACCTTCAAGGTTGGGAGAGAGAG 2811  
Oy AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880  
Db 2812 GATGAGGACCTTTCATCACTGAGCGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2871  
Oy ArgGlnAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlySerPro--- 899  
Db 2872 CCGAGAGATGTGGCGGTGACCACTTGTGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2931  
Oy ProAlaArgThrProMetGlnValThrValThrValLeuAspValIleAsnAspProPro 919  
Db 2932 CCGCTTACCGCGCGTGTGAGAAATCAGGTATCAATCTTGTGACATTAATGACAGCCCG 2991  
Oy ValPheGlnGlnAspGlnPheAspValPheValGlnGlnAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTTTGAAGAGAGCAACCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3051  
Oy ValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGCGCAAGATTCGTGTCAACAGACCTGATGAAGCCCTTAATGCCAGATCATATATCAG 3111

QY 960 ILVAlGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
Db 3112 ATTGTGGAAGGAGGACATGCGCATTTCTTCAGCTGAGCTGGCTCAACGGGGAGCTGGCT 3171  
QY 980 ALaleuValAspLeuAspIleGlyValAspArgProGluValLeuValIleGlnIleThr 999  
Db 3172 GCCATGTGAGTGGAGCTTGTGAGTCCGGCGGAGATGTGCTGGTGGTGGAGCGCA 3231  
QY 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGTGGTGGAGCCAGCCAGTGCATCTTCTCGTGGACCAAGATGACAC 3291  
QY 1020 ProProValLeuGlyAsnPhenGlyIleLeuPheAsnAspIleThrAsnArgSerSer 1039  
Db 3292 CCGCTGTGCTGGCCGACCTTCAGATCTCTTCAACAATATGACCAAGTCCAC 3351  
QY 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGCGTGTATCGGTGCATCCCGCCATGACCCCGACGTGTGACAGC 3411  
QY 1060 LeuThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuAsnAlaSerThr 1079  
Db 3412 CTCACTACACCTTCGTGGAGGGACAGAGCTGGCTGTGGTGGACCCCGCCACG 3471  
QY 1080 GlyGlyLeuLeuYAspSerArgAlaLeuAspAsnAsnArgProLeuGlyAlaIleMetSer 1099  
Db 3472 GCGGAACGTGAGCTCAGCGCCGACCTGGACACACCGCGCGTGGAGGGCTCAAGAG 3531  
QY 1100 ValLeuValSer---AspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThr 1118  
Db 3532 GTGTCTGTGTCTGACATGTGCATCCACAGGTCAACCGCTTGTGACCCCTCGTGCACC 3591  
QY 1119 IleIleThrAspGlnMetLeuThrHisSerIleThrLeuArgLeuGlyAsnMetSerPro 1138  
Db 3592 ATCATACGAGACCAATGCTGACCAACAGCATCTGTCCGCTGGAGAACATGTCCAG 3651  
QY 1139 GlyArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAla 1158  
Db 3652 GAAAGATTCCTGTCCCGCTGTGGCTCTCTGTGGAGGGGGTGGCCGCTGTGTCC 3711  
QY 1159 ThrProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGly 1178  
Db 3712 ACCACCAAGAGCAAGCTTGTGTTCACGTCCAAAGACCAACCAAGCTC---AGCTCC 3768  
QY 1179 HisIleLeuAsnValSerLeuSerValGlyGlnProProGlyGlyProPro 1198  
Db 3769 AACAATCTGAACCTGACCTTCTGGCGCTGCTGCGCGCGCTCCGCGCGC-----CAG 3822  
QY 1199 PheLeuProSerGlyAspLeuGlnGlyArgLeuTyrLeuAsnArgSerLeuLeuThrAla 1218  
Db 3823 TTCCTCCCGTGGAGGACCTGACAGAGCAAGATCTACCTGAATCGAGCGTGTGACACC 3882  
QY 1219 IleSerAlaGlnArgValIleProPheAspAsnIleCysLeuArgGluProCysGlu 1238  
Db 3883 ATCTCCACGAGGCGGTGCTGCTTGCAGACCAACTGCTGCGCGAGCCCTTCAGAG 3942  
QY 1239 AsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSer 1258  
Db 3943 AACTACATGAAGCGGTGCTGCTTCCGATTCGACAGCTCCCGCGCTTCTCACTCC 4002  
QY 1259 SerSerValLeuPheArgProIleHisArgValGlyLeuArgCysArgCysProPro 1278  
Db 4003 ACCACCGTCTCTCCGCGCCATCCACCCATCAACGGCTGCGCTCCGCTCCGCGCC 4062  
QY 1279 GlyPheThrGlyAspTyrCysGlnIleValIleAspLeuCysTyrSerArgProCysGly 1298  
Db 4063 GGGCTTACCGGAGCACTAGCGAGACGAGATGACCTTGTGCTACCTCCAGCCGTGGG 4122  
QY 1299 ProHisGlyArgCysArgSerArgGlyGlyTyrTyrCysLeuCysArgAspGlyTyr 1318  
Db 4123 GCCAAGCGCGCTGCGCAGCCGCGAGGGGCGTCAACCTGCGAAGTGTCTTCAGAGACTTC 4182

QY 1319 ThrGlyGlnHisCysGlyValSerAlaArgSerGlyArgCysThrProGlyValCysIys 1338  
Db 4183 ACTGGAGAGCACTGTGAGGTGATGCCGCTCAGGCGCTGTGGCCAAAGGGGTGGCAG 4242  
QY 1339 AsnGlyGlyThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysProSerGly 1358  
Db 4243 AACGGGGGACCTGTGGAGCACTGTGATCCGCGGCTTCCATGCTGTGTCTCTCCG 4302  
QY 1359 AspPheGlyValProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIle 1378  
Db 4303 GAGTATGAGAGGCGCTTACTGTAGGTGACACAGAGCTTCCCGCCAGTCTTCTGTC 4362  
QY 1379 ThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIys 1398  
Db 4363 ACTTCCGGGGCTGTGAGCAGCGCTTCCATCCATCCATCCCTCAAGTTCCACTCAG 4422  
QY 1399 GlyArgAspGlyLeuLeuLeuTyrAsnGlyValArgPheAsnGlyIleHisAspPheValAla 1418  
Db 4423 GAAAGGAGCGGTGTCTTCTTCAACGCGCTTCAATGAGAGCAAGCATTCATCGCC 4482  
QY 1419 LeuGlyValIleGlnGlyGlnValGlnLeuThrPheSer-----AlaGlyGlySerThr 1436  
Db 4483 CTGAGATGTGTGAGAGAGAGGTGACGTCACTTCTGTGACGTGCGAGCCAGACACA 4542  
QY 1437 ThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGln 1456  
Db 4543 ACAGCCGTGGACCCAGAGGTTCCAGTGTGTGAGTGAAGGGCGGTGGCACTGTGGAG 4602  
QY 1457 LeuIysIleTyrAsnIys-----ProLeu 1464  
Db 4603 GTGCACTACTTACAAAGAGTAAGTGGCCCAACCACTTCCCTGGCCCGCCAGCCCAAT 4662  
QY 1465 LeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnIleValAlaValThrVal 1484  
Db 4663 ATTTGGCACTGGGCTGGCCCATATGGCGCTCCGGGAAAGATGGCCGTGTGACAGTG 4722  
QY 1485 AspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGlyAsnTyrSerCys 1504  
Db 4723 GATGATTGTGACAAACCATGCTGTGCGCTTGTGAAGAGCAATGGGAACTACAGCTGC 4782  
QY 1505 AlaAlaGlnGlyThrGlnGlyGlySerIysIysSerLeuAspLeuThrGlyProLeuLeu 1524  
Db 4783 GCTGCCAGGCACTGACCGGCTTCAGAGATCTCGAATCTGACCGGCGCTTACTC 4842  
QY 1525 LeuGlyGlyValProAspLeuProGlyLeuSerPheProValArgMetArgGlnPheValGly 1544  
Db 4843 CTGGGGGGGTGCTCCCACTGTCAGAGACTTCCCAAGTGCACAAACCGGAGTTCCGTGGC 4902  
QY 1545 CysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAspPheIleAlaAsn 1564  
Db 4903 TGCATGCGGAACTGTCAGTGCACCGCAAAATGTGGCATGTGGCGGATTCATGCCAAC 4962  
QY 1565 AsnGlyThrValProGlyCysProAlaIleValIysValAsnValCysAspSerAsnThrCysHis 1584  
Db 4963 AATGGACCCGGAAAGGCTGCGCTGCGAGAACTTCTGAGTGGAGGCGGTGTCCAG 5022  
QY 1585 AsnGlyGlyThrCysValAsnGlnIleThrAspAlaPheSerCysGlyCysProLeuGlyPhe 1604  
Db 5023 AATGGAGGCACTGTGTCAAGAGTGAATATGTATCTGTGTGAGTGTCACTCCGATTC 5082  
QY 1605 GlyGlyIysSerCysAlaGlnIleMetAlaAspProGlnIlePheLeuGlySerSerLeu 1624  
Db 5083 GCGGGGAAAGATGTGAGAGCAATGCTCCACCCCAAGCTTTCAGCGGTGAGAGCGTC 5142  
QY 1625 ValAlaThrPheIleGlyLeuSerLeuProIleSerGlnProTyrTyrLeuSerLeuMetPhe 1644  
Db 5143 GTGTCTGAGAGCACTTAACATCACTCTGTGGCTTGTGCTGCTGAGGCTCATGTTTC 5202  
QY 1645 ArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIle 1664  
Db 5203 CGGACCGGAGAGAGAGACAGGCTTCTGATGAGGCGCACAGTGTGTGGCCACAGCTTC 5262  
QY 1665 ThrLeuGlnLeuArgGlnGlyHisValIleMetLeuSerValGlnGlyThrGlyLeuGlnAla 1684









Db 2757 TACTCAAGCTGTAACATGATCGACCGGCTTCATGATCACTGCTATGTCAC 2816  
 Qy ValAsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisArgThrVal 719  
 Db 2817 ATCAACATCAAGATGCGCAACATCACTGCGGCTCTTCAAGTGGCCACTACTCAGTG 2876  
 Qy AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2877 AGTGTGAATGAAATCCGGCCATGGGTAGCACATAGTGTCATCACTGCTCTGATGAT 2936  
 Qy AspThrGlyGluAsnAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2937 GACCTGGTGAAGATGCTCGATCACTATCTCTGGAGGACCAACCTGCCCACTTCGCG 2996  
 Qy IleAspAlaAspThrGlyAlaValThrThrGlnIleGluLeuAspArgGlyVal 779  
 Db 2997 ATTGATCAAGATCAAGAGCCATTAACAGGCCCATTAAGCTATGAGACCAAGTG 3056  
 Qy SerThrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIlySerAspThrThr 799  
 Db 3057 ACCTACACCTCGCTATCAACAGCTCGGGACATGGCATCCCAAGAGGACAGACTACT 3116  
 Qy ThrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 3117 TATGTGAGGATGATGTCATATGACGTAAATGACATGCTCCACAAATTTGGCTCCAC 3176  
 Qy TyrGlnIlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 3177 TATACAGGCTGCTCTGAGAGATGCCCACTTTCACAGTGCCTGCAATCTCAGCC 3236  
 Qy ThrAspArgAspSerGlyLeuAsnGlyArgValPheThrThrPheGlnIlyAspAsp 859  
 Db 3237 ACTGACCGGAGTGTCTCATGCCAAATGGCGGGTCAAGTACACTTCCAGAAATGGGAGAT 3296  
 Qy GluAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgAspLeu 879  
 Db 3297 GGGGATGAAGATTTTACCATGAGCCCACTGTGAATGTCCGTAAAGTAAAGGCGGCTA 3356  
 Qy AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIlyGlyMetPro 899  
 Db 3357 GACCGGAGGAGATACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3416  
 Qy ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspPro 919  
 Db 3417 CCACTCCGGAAGTCCAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3476  
 Qy ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 3477 GCTTCCCAAGCTGAGATTTGAGGTGCGGTGAAGAGATATGATGATGATGATGATGATG 3536  
 Qy ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3537 GTGGCCAGATCACTGCAAGTGAACCTGAGAGAGGCCCAATGCGCATTAATATATACAG 3596  
 Qy IleValGluGluAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3597 ATCTGGAGGGGAAATCCCTGAGCTGTCCAAATGACATCTTCTGAGGAATCAAGC 3656  
 Qy AlaLeuValAspLeuAspArgGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 3657 GCATCTATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3716  
 Qy SerAlaProLeuValSerArgAlaThrValHisValArgLeuAspArgAspAsn 1019  
 Db 3717 TCTGCTCTTTGGTCAAGCGGGCCACTGTGCACTCGCGCTGTTGACAGAAATGACAC 3776  
 Qy ProProValLeuGluAsnPheGluIleLeuPheAsnMetTyrValThrAspArgSer 1039  
 Db 3777 AGCCTGTGCTCAACATCTCAAGATCTTCAACAACATATGATCAACCGTTTCAGAC 3836  
 Qy SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 Db 3837 ACCTTCCGTCGGGCAATATTTGGCGCATCCCACTTAATGACCCCGATGTCTCCGACAC 3896

Qy LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 Db 3897 CTCTTACTCTCTTATGAGGTGGCAATGATGATGATGATGATGATGATGATGATGATGAT 3956  
 Qy GlyIleLeuIlyLeuSerArgAlaLeuAspAsnAspArgProLeuGluAlaIleMetSer 1099  
 Db 3957 GGGAGCTCGACTCAGCGGAAAGCTAGCAATATACCGGCCACTGTGCTGCTCAATGTTG 4016  
 Qy ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db 4017 GTACTGTCAAGATGCTGCAACAGTGAAGGCGGAGTGTGTGCTGGCGGTGTATC 4076  
 Qy IleThrAspGluMetLeuThrHisSerIleThrLeuArgGluAspMetSerProGlu 1139  
 Db 4077 ATACGGAGAGATGCTGTGGCAACAGCTGACCGTGTGCTTGAAGAACATGTGGAGAG 4136  
 Qy ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
 Db 4137 CGCTTCTGTCAACCGCTGGGCGCTCTGAGAGGCGGTGCTGCGTGTGCTGCTGCTGCT 4196  
 Qy ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
 Db 4197 CCGCTGAAGAGATCTTCACTTCAACATCAAGAACACAGACAGCTA---GGGGACAC 4253  
 Qy IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyPro----- 1197  
 Db 4254 GTGCTCAATGTAGTTTCTCGGCGTACGCTCAACGTTGAGGGCGGGCGGGCGCTGACAGG 4313  
 Qy ProPheLeuProSerGluAspLeuGlnIlyGlyLeuTyrLeuAsnArgSerLeuLeuThr 1217  
 Db 4314 CCTTGTTGACGCTCCAGAGAGCTGCAAGAGATTTGATGATGATGATGATGATGATGAT 4373  
 Qy AlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCys 1237  
 Db 4374 GCTCGCTCTCTGCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4433  
 Qy GluAsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAla 1257  
 Db 4434 GAAATCAATCAATGAAATGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4493  
 Qy SerSerSerValLeuPheArgProIleHisProValGlyIlyLeuArgCysArgCysPro 1277  
 Db 4494 TCGGCTTCAACGCTGTTCCGACCCATCCAGCCCATGCTGCGCTGCTGCTGCTGCTGCTG 4553  
 Qy ProGluPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCys 1297  
 Db 4554 CCGGATTCACGGAGATCTTTGCGAGACCGAGCTGACCTGCTGCTGCTGCTGCTGCTGCT 4613  
 Qy IleProIleGlyArgCysArgSerArgGluGlyIlyTyrThrCysLeuCysArgAspGly 1317  
 Db 4614 CGAACGGGCGAGCTGCGCGCGCGCGCGAGGGAGGCTACAGCGCTGCGCGCGCGCGCG 4673  
 Qy TyrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCys 1337  
 Db 4674 TTACCGGAGAGGACCTGCAAGCTGACACAGAGCGCGCGCTGCTGCGCGCGCTGCTG 4733  
 Qy LysAsnGlyIlyThrCysValAsnLeuValGlyIlyPheLysCysAspCysProSer 1357  
 Db 4734 CGGACGGGGGACCTGCAACCGAGCGGCCCAAGGGCGCTTTCGCTGCGACGTCGGCA 4793  
 Qy GlyAsp---PheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSer 1376  
 Db 4794 GCGGCGCTTCAAGGGCGCGCGCTGAGAGTGTGCGCGCTGCTTCCGCGCACTTG 4853  
 Qy PheIleThrPheArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAla 1396  
 Db 1377 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4913  
 Qy ThrLeuGluArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIlyAspPhe 1416  
 Db 4914 AAGTGCAGAGAGCGGGGTGCTTCTTCAACGGGCGCTGAACGAGAGACGACTTC 4973

QY	1417	ValAlaLeuGIValAlaLeGIInGIInValAlaLeuThrPheSerAlaGlyLeuSerThr	1436
Db	4974	CTGGCCCTGGAACTCTGGCTGGGCCAAATGGGGCTCAATATTTTCCACGGGGAATCCAAAC	5033
QY	1437	ThrThrValSerProPheValProGIyAlValSerArgGIyInTPHLeThrValGIIn	1456
Db	5034	ACCGGTGCACCCCAAGTTCCAGGGGGCTTGAATGACGGGCATGACATACAGTGCAT	5093
QY	1457	LeuYerTYrAAsnLysProLeuLeuGIyInThrGIyLeuProGIInGIyProSerGIu	1476
Db	5094	CTGAGATCTACAAACAAGCCCCCGGACAGATGCCCTTGAAGGGGTGACAGAGGCCCTCCAAAG	5153
QY	1477	GIInLysValAlaValAlaThrValAlaSerGIyCysAspThrGIyValAlaLeuArgPheGIy	1496
Db	5154	GACAAAGGTGGCTGGCTGAAGGGGTGATATTTGATATGGCCCGTGGCTCTCCAGATTGGT	5213
QY	1497	SerValLeuGIyAAsnTYrSerCysAlaAlaGIInGIyThrGIInGIyGIySerLysLysSer	1516
Db	5214	CGTGAAGATGGCACTACTACATGCCCGGCTGGTGTGGCAACAAGCTCCAAAGAATCC	5273
QY	1517	LeuAspLeuThrGIyProLeuLeuLeuGIyGIyValProAspLeuProGIuSerPhePro	1536
Db	5274	CTGACACCTGAAGGGGCCCTCTTCTTCTGGAGAGGTGCCCAACCTCCCGAACAATCCCC	5333
QY	1537	ValArgMetArgGIInPheValGIyCysMetArgAsnLeuGIInValAspSerArgHisIle	1556
Db	5334	GTATCCCAATGAAGCTTCATCGGCTGTATGGGGACCTGCACATGATATGGCCGCGCAATGG	5393
QY	1557	AspMetAlaAspPheIleAlaAsnAsnGIyThrValProGIyCysProAlaLysLysAsn	1576
Db	5394	GACATGGGGCTTTGTGTGGCAATATATGGACCATGGCAGGCTGCAAGCCCAACTACAC	5453
QY	1577	ValCysAspSerAsnThrCysHisAsnGIyGIyThrCysValaAsnGIInTPRAspAlaPhe	1596
Db	5454	TTTTGTGACTCAGGGCCCTGCAGAAAGACGTGGCTTGCCTGAGAGCGCTGGGGACAGTTC	5513
QY	1597	SerCysGIuGIyProLeuGIyPheGIyGIyLysSerCysAlaGIInGIuMetAlaAsnPro	1616
Db	5514	AGCTGCACCTGCCCTTGGGGCTTGGCGGCAAAACATGTCTACGCTTATGGCCCAATCCC	5573
QY	1617	GIInAspPheLeuGIySerSerLeuValAlaTPHHis---GIyLysSerLeuProIleSer	1635
Db	5574	CACCATTTCCGTGGCAACGGACACACATGACGTGAACCTTGGAAATGACATGGCTGTGCT	5633
QY	1636	GIInProTPrGIyLeuSerLeuMetPheArgThrArgGIInAlaAspGIyValLeuLeuGIIn	1655
Db	5634	GTGSCATGGTACCTGGGGCTGGCATTTCCAGACAGGGGAACGACGGGGGTCTGATGCCAA	5693
QY	1656	AlaIleThrArgGIyLysSerThrTleThrLeuGIInLeuArgGIyInHisValMetLeu	1675
Db	5694	GTGCAAGCTGGGACACACAGACACCTCTTTGACAGCTAAGATCGGGGGTATCTGTCTGG	5753
QY	1676	SerVal---GIuGIyThrGIyLeuGIInAlaSerSerLeuAspLeuGIuProGIyArgAla	1694
Db	5754	ACAGGTACACAGGGGCTGGGCT---CGTGGTCCCAATCTCTTCTTGACACAGGTACATGCTC	5810
QY	1695	AsnAspArgIyAspTPRHisHisAlaGIInLeuAlaLeu-----GIyAlaSerGIyGIyPro	1712
Db	5811	AGTATAGGCGCGGTGGACAGATCTGGCGCTGGAGTTGCAGAGAAACCAAGGTGGCGGGCGG	5870
QY	1713	GIyHisValaIleLeu-----SerPheAspTYrGIyGIInIleArgAlaGIyGIyAsnLeu	1730
Db	5871	GGCCACACATGCTTATAGTGTCTACCTGACCTTTCAGCTTCCAGACACCATGGCGGTG	5930
QY	1731	GIyProArgLeuHisAspGIyLeuHisLeuSerAsnIleThrValGIyGIyIle---ProGIy	1749
Db	5931	GGAGATAGCTGCAGGGCTGAAGGTAAAGACATGCCACGTGGAGAGGCTGGCCCCGGGC	5990
QY	1750	ProAlaGIyGIyValAlaAlaArgGIyPheArgGIyCysLeuGIInGIyValArgValSerAsp	1769
Db	5991	AGTGCACAGAGAGGCTCTCAGAGGTCTGGTTGGCTGCATCCAGGGGGTGTGGCTGGCTCC	6050
QY	1770	ThrProGIuGIyValaAsnSer---LeuAspProSerHisAspGIyLysIleAsnValGIu	1788

Db	6051	ACACCTCTGGAGCTCCCGGCCCCCTGACACCCCCACAC-----CGAGTAAATCGGAG	6104
Qy	1789	GIINGLYCYSSerLeuProAspProCYsaAspSerAsnProCYsProAlaAsnSerTYrCYs	1808
Db	6105	CTGGCTGTGTGTGACCAACAGCCGTGCTCTGGGCGCTCGCCACCTCACGACAGATGCG	6164
Qy	1809	SerAsnAspTPraAspSerTYrSerCYsSerCYsaAspProGLYTYrTYrGLYAspAsnCYs	1822
Db	6165	CGGAGCTCTGTGACAGACTTTTCTTGACCTCGCACCGAGGTACTATACGGCCACAGCTGT	6222
Qy	1829	ThraAsnValCYsaAspLeuAsnProCYsGLuHisGLuInsValCYsThrArgLysProSer	1848
Db	6225	GTGAGTGCCTGCCTCTTAACCCCTGTGTACAGAACAGAGGATCATGCGCGACCTGCAAGA	6288
Qy	1849	AlaProHisGLYTYrThrCYsGLuCYsProProAsnTYrLeuGLYProTYrCYsGLuThr	1866
Db	6285	GGCCCCACATGGCTATACCTGTGACCTGTGTGGGTGGCTATTTCGGGACACACTGTGAGAC	6344
Qy	1869	ArgLLeaAspGLnProCYsProArgGLYTYrTProGLYHisProThrCYsGLYProCYsaAsn	1888
Db	6345	AGGATGACACAGACAGTCCACCGGGGCTGTGGGGAGCCACCACTGTGGCCCCCTGACAC	6404
Qy	1889	CysAspValSerLysGLYpHeaAspProAspCYsaAsnLysThrSerGLYGLuCYsHisCYs	1908
Db	6405	TGTGTATGTTCCAAAGGTTTGTGATCCCAACGCAACAAGCAAAATGGGCAAGTCACTGC	6464
Qy	1909	LysGLuAsnHisTYrArgProProGLYSerProThrCYsLeuLeuCYsaAspCYTYrPro	1928
Db	6465	AAGAGTTTCCACTACAGACCGGGGACAGTACTTTCCTCCCATGATGACTGTCACTCT	6524
Qy	1929	ThrGLYSerLeuSerArgValCYsaAspProGLuAspGLYGLuCYsProCYsLysProGLY	1948
Db	6525	GTGGGCTTCCACTCGCGCTCATGTGACACCCACAGCGGGGAGTGCGCCCTGTGCGCCAGGA	6588
Qy	1949	ValLleGLYArgGLuCYsaAspArgCYsaAspAsnProPheAlaGLuValThrThraAsnGLY	1966
Db	6585	GGCTTGGCGCGCAGTGCACAGCAGTGTGACAGTCCCTTCGACAGAGTGACAGCGACGGC	6644
Qy	1969	CysGLuValAsnTYrAspSerCYsProArgAlaLleGLuAlaGLYTYrTProProArg	1988
Db	6645	TGCGGGGTGCTATGATGATGCTGTGCTTAAGTCCCTGAGATCTGGTGTGTGTGGCCCCAG	6704
Qy	1989	ThrArgPheGLYLeuProAlaAlaAlaProCYsProLysGLYSerPheGLYThrAlaVal	2008
Db	6705	ACAAAGTTTGGCGTCCGCGCACAGTGCCTCTGTGCCCGGGGGCGCTGTGGGTGCTGTGTG	6764
Qy	2009	ArgHisCYsaAspGLuHisArgGLYTYrPheProProAsnLeuPheAsnCYsThrSerLle	2028
Db	6765	CGGCTGTGTATGAGGCCACAGGTGGTGTGAGGCCACCTCTTCACATGTACTCTCCCT	6824
Qy	2029	ThrPheSerGLuLeuLysGLYpHeaLleGLuArgLeuGLuArgAsnGLuSerGLYLeuAsp	2048
Db	6825	GCTTTTGAAGAGCTCAGTGTGCTGTGATGGCTTAAGCTGAACAGACAGGCACTGGAT	6884
Qy	2049	SerGLYArgSerGLuInsLleAlaLeuLeuLeuArgAsnAlaThrGLuHisThrAlaGLY	2068
Db	6885	ACCAATGAGGCAAGAAAGTGGCTCACAGCGGCTACGGAGGTGACTGGCCACACATGACAC	6944
Qy	2069	TYrPheGLYSerAspValLysValAlaLysTYrGLuInsLleAlaThrArgLeuLeuAlaHisGLY	2088
Db	6945	TATTTTGAAGCAAGATGTTCAAGTCACTGCCCGCTGTGGCCACCTGTGGCTTGTGAG	7004
Qy	2089	SerThrGLuArgGLYpHeaGLYLeuSerAlaThrGLuAspValHisPheThrGLuAsnLeu	2108
Db	7005	AGCATATCAGCAGGGCTTCGGCTGTACAGGCCACACAGAGATGCCCACTTCATATGAATACTGT	7064
Qy	2109	LeuArgValGLYSerAlaLeuLeuAspThrAlaAsnLysArgHisTYrGLuInsLle---	2127
Db	7065	CTGTGGGCGCGCTGTGACTGTGCTTGGCCCCAGACAGGGGACTTGTGGCGGCGCTGGGG	7124
Qy	2128	GIINGLThrGLuGLY-----GLYThrAlaTPheLeuGLuHisTYrGLuAlaTYr	2144

Db 7125 CAGCGGCGCCCTGGGGCTCCCGACGCGGAGCTGGTAGGCACTGAGAGTAT 7184  
Qy 2145 AlAserAlaLeuAlaGlnMetArgHisThrTyrLeuSerProPheThrIleValThr 2164  
Db 7185 GAGGCCACACTCCCGAGGAATATGGAATCACTGAAATCCCATGGGGCTGGTAGC 7244  
Qy 2165 ProAsnIleValIleSerValValArgLeuAspIleGlyAsnPhe-----AlaGlyAla 2182  
Db 7245 CCAAAATATCATGCTCAGCATTCAGCATGAGCAACCCAGTTCTCCCGGGGGCCCGT 7204  
Qy 2183 LysLeuProArgTyrGluAla-----LeuArgGlyGluGlnProProAsnLeuGluThr 2200  
Db 7305 GCGTACCTGGCTTACCATACCAACTCTTTCAGAGCCAGGATCCCTGGATCTCAACCC 7364  
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
Db 7365 CATGTGTGTGCTTCCAGTCCCGACGCGCATCCCATCTGAAGTTCTGCCACAGAC 7424  
Qy 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgIleArgHisProGluLeu 2240  
Db 7425 AGCAGCATGAGAAATCTCCACACCTCAAGTGTGTCCTCCACACGCGCCGACAGCCA 7484  
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db 7485 GAGCCTGGAGTCTCATTAATCATCTCTCGTTTACCGACCTTAGGGGACTGCTCCCT 7544  
Qy 2261 HisAsnTyrAspProAspIleArgSerLeuArgValProIleAsnProIleAsnThr 2280  
Db 7545 GCCAGTTCAGGACAGAACCGGAGTCCAGGCTTCTCAAGAACCCCGCATGAATCC 7604  
Qy 2281 ProValIleSerIleSerValHisAspAspGluGluLeuProArgAlaLeuAspIle 2300  
Db 7605 CCGGTGTACGCGTGTGTGTTCACAGCAGCACTTCTAAGGGGATCTCGAGTCC 7664  
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgHisProIleCysVal 2320  
Db 7665 CCATACGCTTAAGTTCGCTGCTGCTACAGACCGAATCGAGAGCGGACTGTGTG 7724  
Qy 2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGlyCysGlu 2340  
Db 7725 CAGTGGAGCCACCTGGCTGGCGGAGCAGCATGCTGTGTGACAGACGAGACTGGAG 7784  
Qy 2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisSerThrSerPheAla 2360  
Db 7785 CTGTGTGACAGGATGGGTCCACGACGCGTGTGCTGCGACGCGACGAGCACTTGGG 7844  
Qy 2361 ValLeuMetAspValSerArgArgGlu-----AsnGlyGluIleLeuProLeuGlyThr 2378  
Db 7845 GTCTCATGATGCTCTCCCGTGAAGAGCTGAGAGGACCACTGAGCTGTGGCTGTG 7904  
Qy 2379 LeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeu 2398  
Db 7905 TTCACCCAGCTGGCGGT 7964  
Qy 2399 ThrLeuLeuArgIleLeuArgSerArgGlnHisGlyIleArgArgAsnLeuThrAlaAla 2418  
Db 7965 CTGAGCTGGCGAGCTCAAGTCCAAATGTCGTGGATTCATGCGCAATGTGGACGCGCC 8024  
Qy 2419 LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAla 2438  
Db 8025 CTGGGGGTGGCAGAGCTCTCTCTCTCTCTGGGAAATCAAGACCAATCAGCTGGTG 8084  
Qy 2439 CysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeu 2458  
Db 8085 TGCACTGACAGTGGCATCTCTGCACTACTTCTTCTCAGCACTTCGCGTGGCTTTC 8144  
Qy 2459 LeuGluAlaLeuHisLeuTyrArgAlaLeuThrGlyValAlaArgAspValAsnThrGlyPro 2478  
Db 8145 GTGAGAGGGGTGACCTTACCGCATGCAAGTTGAGCAGACGAGACCGCGGGCC 8204  
Qy 2479 MetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaVal 2498  
Db 8205 ATGGCTTCTACCATGCGCTGGCTGGGCTGCTCTCTGTGTGTGTGTGTGTGTGTGT 8264

Qy 2499 GlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThr 2518  
Db 8265 GCGCTGGACCTCGAGGGCTATGGAACTTAATCTTGTGTGATCTCACTCAAGACCC 8324  
Qy 2519 LeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyr 2538  
Db 8325 CTCATCTGGAGCTTGTGGCCCTGTGCTGTGCTGATGATGAACGGACATGTCTT 8384  
Qy 2539 IleLeuAlaAlaArgAlaSerCysAla-----AlaGlnArgGlnGlyPheGluValValGly 2557  
Db 8385 CTCTGGCTGGCCCGACATCTCTCTCAGAGGCGAGAGAGGCC-----AAGAGACC 8438  
Qy 2558 ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeu 2577  
Db 8439 TCTGACATGACCTTGCGAGCTCTTCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 8498  
Qy 2578 LeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCys 2597  
Db 8499 TTTGGGCTCTGGCAGTCAACACAGCATCTTACGCTTCACTCATCTCATCTGAGCTC 8558  
Qy 2598 AsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGluValArg 2617  
Db 8559 TCGGCTTCAGAGGCTGGCGGTGCTGTGCTTCTGTGTCTTAAATGACAGATCTCGG 8618  
Qy 2618 LysAlaLeuLysLeuAlaCys-----SerArgLysProSerProAspProAlaLeuThrThr 2636  
Db 8619 GGTGCTGTGATCCAGGCTGTGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8678  
Qy 2637 LysSerThrLeuThrSerSerTyrAsnCyAspSerProTyrAlaAspGlyArgLeuTyr 2656  
Db 8679 CTTGGGCTGGGACCTGGGGCTTCAACAACAGCGCTCTTGTGAAGAGTGGCTCAATC 8738  
Qy 2657 Gln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSer 2675  
Db 8739 CCGATCATCTGGAGGCTTCAACCGTCTCTGTGAGAGAGTCCGCTCCGAGCGAGCC 8798  
Qy 2676 GlnPro-----SerTyrIle-----ProPheLeuLeuArgGlu 2686  
Db 8799 CAGACACAGACAGCAGCGGGGCGGAGCTACCTCAGAGGACAAATGCTGTGTGACAT 8858  
Qy 2687 GluSerAla-----LeuAsnProGlyGlnGlyProProGlyLeu 2699  
Db 8859 GGCCTACGCGCTGACCACTGACCAAGCTCTCAGAGCTCATGTCGGCCCACTGACCTG 8918  
Qy 2700 Gly-----AspProGlySerLeuPheLeuGluGlyGlnAspGlnGln 2713  
Db 8919 GAGTGGCCATGTTCCATGAGATCTGGCGCA----- 8951  
Qy 2714 HisAspProAspThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerIleSer 2733  
Db 8952 -----GACTCCGACTCTGACAGTACCTGTCTTGGAGAGAGAGAGTCTCTCC 9002  
Qy 2734 TyrAlaSerThrHisSerSerAspSerGluGluGluGluGluGluGluGluAla 2753  
Db 9003 ATTCATCTTGAGAAAGCAGGACGAGGCGGAGCGGCGCTTCCAGGCCCATCTC 9062  
Qy 2754 AlaPheProGlyGluGlnGlyTyrAspSerLeuLeuGlyProGlyValaGluArgLeuPro 2773  
Db 9063 TGGCCAGCAGCCAG-----AGTGAAGGCTCTTC 9092  
Qy 2774 LeuHisSerThrProLysAspGlyGlyProGlyLysValaProTrpProGly--- 2792  
Db 9093 ACCCAC-----CCCAAGATGTGATGAGCATGACTCTCTCTCACTGCGCAGCCCTG 9146  
Qy 2793 -----AspPheGlyThr----- 2796  
Db 9147 GGGAGCTGCGAGGCGAGCCCTGTGCTGTGCAAGTCTGGGGCTCTGAAGAGCGCTGGGG 9206  
Qy 2797 -----ThrAlaGlyGluSerSerGlyAsnGlyAlaProGluGluLysArgLeuArgLysAsn 2814  
Db 9207 CTGACACACAGAGATGACACTTAACAACAACAGCAGACAGACCGGCGCTG---ACCAAT 9263



QY 2815 G1yAspAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySerSerAlaGlnPro 2834  
 Db 9264 GGGGAT-----GAGACTTCTG-----GGCCGCGCCAGCCGCGAG 9299  
 QY 2835 H1eLyG1y1LeuYslySlySlySlySlySlySlySlySlySlySlySlySlySlySly 2853  
 Db 9300 AGGAAAGGATCTCTGAAAGAACCGTTGCAATACCCCTGGTGCACAGACCGAGGTGCC 9359  
 QY 2854 LeuArgLeuProLeuGluGlnGlyGlnGlySerSerArgGlySerSerAla----- 2870  
 Db 9360 CCTGAG-----CTGCTCGTGTCCGTCGACGACCTTGCGGCGACCGTGACGACGCT 9413  
 QY 2871 -----SerGluGlySerArgGlyGlyProProProArgProProPro 2884  
 Db 9414 GCCTTACGTCGTCGATCTATGCTGCGCGGCGACGCGGACGCTTTCACGACGACGAC 9473  
 QY 2885 ArgGlnSerLeuGlnGluGlnLeuAen 2893  
 Db 9474 CGCTACTCTTCTAGAGAACGCTGAC 9500

## RESULT 14

US-10-149-819-42  
 / Sequence 42, Application US/10149819  
 / Publication No. US20030044913A1  
 / GENERAL INFORMATION:  
 / APPLICANT: INCYTE GENOMICS, INC.  
 / APPLICANT: YUE, Henry  
 / APPLICANT: AZIMZAI, Yalda  
 / APPLICANT: TANG, Y. Tom  
 / APPLICANT: PATTERSON, Chandra  
 / APPLICANT: BAUGHN, Mariah R.  
 / APPLICANT: LU, Dying Aina M.  
 / APPLICANT: SHAH, Preeti  
 / APPLICANT: LAU, Preeti  
 / APPLICANT: AU-YOUNG, Janice  
 / APPLICANT: BURROD, Neil  
 / TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
 / FILE REFERENCE: PR-0760 PCT  
 / CURRENT APPLICATION NUMBER: US/10/149,819  
 / CURRENT FILING DATE: 2002-06-10  
 / PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354  
 / PRIOR FILING DATE: 1999-12-10; 1999-12-16  
 / NUMBER OF SEQ ID NOS: 42  
 / SOFTWARE: PERL Program  
 / SEQ ID NO 42  
 / LENGTH: 11648  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / NAME/KEY: misc feature  
 / OTHER INFORMATION: Incyte ID No. US20030044913A1 2847752CBI  
 / US-10-149-819-42

## Alignment Scores:

Pred. No.: 0 Length: 11648  
 Score: 7766.00 Matches: 1561  
 Best Local Similarity: 66.22% Conservative: 421  
 Best Local Similarity: 52.16% Mismatches: 812  
 Query Match: 49.96% Indels: 200  
 Gaps: 15 Gaps: 49

US-09-916-849A-3 (1-2923) x US-10-149-819-42 (1-11648)

QY 28 ProteuLeuGlyAspGln-----ValGlyProCysArg 38  
 Db 416 CCTCTTGGCGGCGACAGGACCTTTCAGAAAGATGCTGTACACAGGGGCTGTCT 475  
 QY 39 SerLeuGlySerArgGlyArgGlySerSer----- 48  
 Db 476 CAG-----GGGTCCCGGCTCGGAGAACGCTCCGCTCTTCAGACTTTTGATTCGAC 532  
 QY 49 -----G1yAlaCysAlaProMetGlyTTPLeuCysProSerSerAlaSerAen 64

Db 533 ACCAGGTCCTCCAAAGCCGGTCTCTCCAGCGGACGTCGGACAGAGCTCCGCAAAAG 592  
 QY 65 LeuTTPLeuYTrThSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84  
 Db 593 ---TGGCAACCGCGGCTCTGTGGGAAATTATGGCAACG-----GAGCAAGGCT 643  
 QY 85 -----ProHisLeuAspGlyLeuArgValTTPCysProGluSerGluHis 100  
 Db 644 AGGGCGAAGAGACAGACAT-----CCGAGGCAAGAAAGACAG 682  
 QY 101 ILePro-----LeuProProAlaProGluGlyCys 110  
 Db 683 CCCCCCGGGAAGCTGTCTCCAGGGGCTCGGAGTCTGGCCCGGCTG--GATTACAGA 741  
 QY 111 ProTTPSerCysArgLeuLeuGly1LeGlyGlyHisLeuSerProGlnGlyLysLeuThr 130  
 Db 742 CCAAGCAGCGGAGAGACAGCTCTGATCAGAGTTA---GCACCCCGAGCTGTGGACA 798  
 QY 131 LeuProGluGluHisProCysLeuLeuYsAlaProArg-----LeuArgCysGlnSerCys 148  
 Db 799 GCTCCCGAGCGGCGCCAGAGCGCATGCGCTCCGGGGTCTTCCGCTGCGCTTC--- 855  
 QY 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLysArg 161  
 Db 856 ---CTCCGCAAGCGCCCGCGCGCTCCCGGAGCTCCCGCGCTGAAAGCCAG 912  
 QY 162 SerProGluGluSerLeuGlyGlyArgArgLysArgAsnValAsnThrAlaProGlnPhe 181  
 Db 913 AAAGTAACCTGGCGCAACCGGCGACAGCTTGTGTCGCGCGCAACCGCGACCGGAGTT 972  
 QY 182 GlnProProSerTyrGlnAlaThrValProGluGlnProAlaGlyThrProValAla 201  
 Db 973 CCGCAGTCAACATCAACACCTGCTGCGGAGATGAGCAGCAGCAGCGCGGTGCTA 1032  
 QY 202 SerLeuArgAla1LeuProAspGlnGlyGluAlaGlyArgGluGluGlyThrMetAsp 221  
 Db 1033 CGCGGTGCTCAGAGACCGGAGCGCGCGAGCGCGCGCTTACTGCTGTCGCG 1092  
 QY 222 AlaLeuPheAspSerArgSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAlaVal 241  
 Db 1093 GCACTCATGAACAGCGCTGCTGCTGAGCTGTTCAGATCGACCGCGAAGCGCTTATC 1152  
 QY 242 ThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261  
 Db 1153 CGTACGGGCGAGCTCTGAGCCGCGAGCAGTGAAGCTCATCTCTGCTGTGACCGCG 1212  
 QY 262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrLeuValThrAsp 281  
 Db 1213 CAGACACCGGCTGCGCGCTCTCGGCACACAGATGTGGCCGTACAGATACCGCAC 1272  
 QY 282 ThrAsnAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGluAen 301  
 Db 1273 CGCAACGACCACTGCGCGCTTTTGAAGCAAGCGATACCGGAGACCTTCGGAAGAT 1332  
 QY 302 LeuGluValGlyTyrGluValLeuThrValAlaArgAlaThrAspGlyAspAlaProProben 321  
 Db 1333 GTGAGAGAGGCTACCTCATCTGAGCTGCTGCACTGACGCGGACCGCCCGCCAAC 1392  
 QY 322 AlaAsnLeuLeuTyrArgLeuLeu-----GluGlySerArgLysSerProSerGluVal 339  
 Db 1393 GCGAACCTGCGCTACCGCTTGTGGGCGCCGACGTCGCGCGCTGCGCGCGCC 1452  
 QY 340 PheGluLeuAspProArgSerGlyVal1LeuArgThrArgGlyProValAspArgGluGlu 359  
 Db 1453 TTCAGATTAATTCACGCTCGCGCTTATCAGACAGCGCGCGAGTGAGACCGGAGAC 1512  
 QY 360 ValGluSerTyrGlnLeuThrValGluAlaSerArgGlnGlyArgAspProGlyProArg 379  
 Db 1513 ATGAAAGCTATAGCTGTGTGTGAAGCCAGCACAGGCGCCAGAAACCGGCGCGGC 1572  
 QY 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPhe 399  
 Db 1573 TCGGCACTGTGCGCGTACATTAATCTGTCTACAGAAAGACAGATGCTCTCAAGTT 1632

QY 400 SerGluValArgTyrValValGlnValArgGluAspValThrProGluValAlaProValLeu 419  
 DB 1633 AGCGAAGAGCGCTACGGCGGAGGTGCGGAGGATGTCGCCCCCAACAGTCGTGCTG 1692  
 QY 420 ArgValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHisTyrSerIleMet 439  
 DB 1693 CGGTCACCGGCACTGACCGGAGCAAGCAAGCAAGATGGTGGCACTTCAACATCATC 1752  
 QY 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
 DB 1753 AGTGGCAATAGCCGTGACACTTGGCCATGCAAGCTCATCTGGCAGATCCAGGTGTG 1812  
 QY 460 SerProLeuAspTyrGlnThrThrTyrGlyTyrThrLeuArgValAlaArgAlaGlnAspGly 479  
 DB 1813 GCACCTTGCACTTCGAGGAGAGAGAGATATGCTTGGCATCAGGGCGGAGATGCT 1872  
 QY 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspAlaAsn 499  
 DB 1873 GGCCTGGCAGCGCTGTCACACACGCGCTGGCCGACATCCAGTGTGGACATCAT 1932  
 QY 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValPro 519  
 DB 1933 GACCACTTCCTATTTTGTTCAGACGCGCTTCCAACTTCTGTCTTGGAAATGCTCCC 1992  
 QY 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 DB 1993 TTGGGTCACTCAGCATCCATTCATTCAGGAGTCATGACATCCATGGGGAAATGCGACA 2052  
 QY 540 LeuGlnTyrArgLeuAlaGlyValGlnHisAspPheProPheThrIleAsnAsnGlyThr 559  
 DB 2053 TTGGAGTACCTCCCTACTGATGTGGCACTGATCATCTCTTTGTGTAACACCGGCAC 2112  
 QY 560 GlyTyrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 DB 2113 GGTGGGTCTCTGAGAGTGTCCTCCCTGGACCGTGAAGTCTGGAGCATTACTCTTTGGT 2172  
 QY 580 ValGlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 DB 2173 GTGAGGCTCGAAGACCAATGCTCAACCCCACTCTGCTCAGCGATGTCACCGTGACT 2232  
 QY 600 ValLeuAspValAlaAspAsnAsnProThrPheThrGlnProGluTyrThrAlaArgLeu 619  
 DB 2233 GTCTGGACCTTAATGACATCGGCTGAGTTCAATGAAGAGTACCACTTACAGCTG 2292  
 QY 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAlaAspArgAspAla 639  
 DB 2293 AATGAGATGCACTGTGGGCAACAGTGTGTGAGTGACCGAGTAAAGCCGTGATGCC 2352  
 QY 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 DB 2353 AACAGTCCCAATCAGATCAAGATCAAGCGGCAACACCGGAATCGCTTGGCATCAGC 2412  
 QY 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnArg 679  
 DB 2413 ACCGAGGGGAGTGTGGGTGTGGATCTGTGCTGTGCACTGAGCTACAGAGAGAGCG 2472  
 QY 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
 DB 2473 TACTTCAAGCTGGAATCAATGATGATGACCGTGGCTTCAATGATCACTGATGAC 2532  
 QY 700 ValAsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrVal 719  
 DB 2533 ATCAACATCAAGATGCAACATCATCGGCGGTCTTTCMAAGTGGCCCACTCATGAG 2592  
 QY 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 DB 2593 AGCTGTGAATGAAGTGGCCCAATGGGTAGCACTAATGTCTCATCACTGCTGTGATGAT 2652  
 QY 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 DB 2653 GACGTGGGTAGAATGCTCGATCACTATCTCTGAGAGACAACCTGCGCCAGTTCCGC 2712

QY 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 DB 2713 ATTTGATGACAGACTGAGAGCCATTACACAGGCCCATTAAGACTATAGAGACAGAGGTG 2772  
 QY 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThr 799  
 DB 2773 ACCATCACTTGGCTATTCACAGCTGGGACAAATGGATGCCACAGAGGACAGACTACT 2832  
 QY 800 TyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 DB 2833 TATGTGGAGGTATGTGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2892  
 QY 820 TyrGlnGlySerValTyrGluAspValProPheThrSerValLeuGlnIleSerAla 839  
 DB 2893 TATACAGGGCTGTGCTGAGAGATGCCCACTTTCACAGATGTCTGTGAGATTCACGCC 2952  
 QY 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAsp 859  
 DB 2953 ACTGACCGGAGATGCTCAGTCCATGCGCGGAGTCCAGTACATCTTCCAGATGTGAAGT 3012  
 QY 860 GlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeu 879  
 DB 3013 GGGAGATGAGATTTTACATGATGAGCCACCTTGGAATGTCCGTACAGTAAAGCGGCTA 3072  
 QY 880 AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetPro 899  
 DB 3073 GACCGGAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3132  
 QY 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnPro 919  
 DB 3133 CAATCCGAGCTCCAGTCAATGATCCAGATGATGATGATGATGATGATGATGATGAT 3192  
 QY 920 ValPheGlnGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 DB 3193 GTCTTCCAGCTGAGAGAGATTTGAGGTGCGGGTGAAGAATATGATGTGGGCTCAGTG 3252  
 QY 940 ValAlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyrGln 959  
 DB 3253 GTGGCCCAAGTACATGAGTGAAGCCCTGACAAAGGCCCAAGCCCATTAATATGACAG 3312  
 QY 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 DB 3313 ATCTGAGAGGGAACATCTCTGATGCTTCCAAATGACATTTCTCTGAGAACTGACG 3372  
 QY 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 DB 3373 GCACTCATGACCTTAATGATGAGCTGCGCAAGATATGATGATGATGATGATGATGAT 3432  
 QY 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 DB 3433 TCTGCTCTTTGGTCAAGCGGCGCACTGTGCAAGTCCGCTGTGATGACCAAGATACAC 3492  
 QY 1020 ProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
 DB 3493 AGCCCTGTCTCAACAATCTTCAGATCTCTTTCACCACTAATGATTCACACCGTTCAAGC 3552  
 QY 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 DB 3553 ACCCTCCGTCGGGATTAATGGGCGCATGCCAGCTTATGACCCGAGATGTCTCCACAC 3612  
 QY 1060 LeuThrTyrSerPheGluArgGlyValAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 DB 3613 CTCTTCTACTCTTTGAGGTGGCAATGATGCTGCAAGCTGTGTGATGACCAACAGAGCT 3672  
 QY 1080 GlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluValAlaIleMetSer 1099  
 DB 3673 GGGAGCTGCACTGACCGGAAAGCTAAGACATTAACCGCCACTGTGTGCTCCATGTTG 3732  
 QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 DB 3733 GTGAGCTGCACAGATGGCTGCAAGCGTGAAGCGGAGTGTGTGTGCTGCTGTGATC 3792  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139

```

Db      3793 ATACGAGGAGGTGCGGCCAACAGCTGACCGTGGCTTGAACAATGTGGCAGAG 3852
Qy      1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnIleValAlaIleThrLeuAlaThr 1159
Db      3853 CGCTTCCTGACACCGCTGCGGCGCTTCTCGAGGGCGGTGGTGGCTCGCTGCTAG 3912
Qy      1160 ProProAerPheIleValValPheAerValGlnAerPheThrAerAlaProGlyGlyIle 1179
Db      3913 CCGCGTAGAGGAGCTTCACTTCAACATCCAGAACAGACAGCGA---GGGGGACCC 3969
Qy      1180 IleLeuAerValSerLeuSerValGlyGlnProProGlyProGlyGlyPro----- 1197
Db      3970 GTGCTCAATGTAGATTCTCGGGCTAGCTCAAGTGGGGCCGGGGCGCGCTGGACGG 4029
Qy      1198 ProPheLeuProSerGluAerPheGlnGlnAerGlyLeuThrLeuAerPheSerLeuThr 1217
Db      4030 CCGTGGTTACGCTCCAGAGAGCTGCAGAGACAGTTGATCGTGGCGGGGGCGCTGGCG 4089
Qy      1218 AlaIleSerAlaGlnAerValLeuProPheAerAerAerIleCysLeuAerGluProCys 1237
Db      4090 GCTCGCTCCCTGCTGCAGACTGCTGCTTGCAGACAGAGTGTGCTGGAGAGCCCTGT 4149
Qy      1238 GluAerPheMetArgCysValSerValLeuAerPheAerSerAlaProPheIleAla 1257
Db      4150 GAGAACTACATGAATATGCGTCCGCTCCGCTTGACTGTCGCGCTTCTGCGCC 4209
Qy      1258 SerSerSerValLeuPheAerProIleAerProValGlyGlyLeuAerCysAerCysPro 1277
Db      4210 TCGGCTTCCAGCTGTTCCAGACCAATCCAGCCAGCTGCGCTGGCGCTGGCGCTCCG 4269
Qy      1278 ProGlyPheThrGlyAerPheCysGlnThrGlnValAerLeuCysAerSerAerProCys 1297
Db      4270 CCGGAGATTCACGGGAGACTTTTGCAGACCGAGCTGAGCTGCTGCTCAACCAATGT 4329
Qy      1298 GlyProIleGlyAerCysAerSerAerGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1317
Db      4330 CCGAACGCGAGAGCTGCGCGCGCGCGAGAGAGCTACACGTCGCTGCGCGCGCGCG 4389
Qy      1318 TGTThrGlyGlyIleAerCysGlnValSerAlaAerSerGlyAerCysThrProGlyValCys 1337
Db      4390 TTCAACGGAGAGAGACTGCGAGAGCTGCAGACCGAGAGCGCGCGCTGCGCGCGCTCC 4449
Qy      1338 LysAerGlyGlyThrCysValAerLeuLeuValGlyGlyPheLeuCysAerCysProSer 1357
Db      4450 CCGAACGCGGGGCACTGCACCGACCGCGCCCAAGCGGCTTTGCTGCGACAGTCCGGGA 4509
Qy      1358 GlyAer---PheGlyLeuProGlyCysGlnValThrThrAerSerPheAlaIleSer 1376
Db      4510 GCGGCGGCTTGCAGGGCGCGCGCTGCGAGGTGGCTGCGCGCTTCCCGCCAGTTG 4569
Qy      1377 PheIleThrPheAerGlyLeuAerGlnAerPheIlePheThrLeuAlaLeuSerPheAla 1396
Db      4570 TTGCTCATGTTTGGCGGCTGCGGCGAGCATTCACCTTACGCTGCTGCTGCTGCGG 4629
Qy      1397 ThrLysGlnAerAerPheLeuLeuLeuGlyAerGlyAerPheAerGlyAerPheAer 1416
Db      4630 ACGAGTSCACAGAGCGGGCTGCTTCTCAACGGGCGCTGAACAGAGACAGCACTTC 4689
Qy      1417 ValAlaLeuGlnValIleGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThr 1436
Db      4690 CTGGCCCTGAACTCTGGCTGGCCAAAGTGGCTCAATATTCACCGGTGAATCCAAC 4749
Qy      1437 ThrThrValSerProPheValProGlyGlyValSerAerPheGlyGlyIleThrValGln 1456
Db      4750 ACCGTGTGACGCCCAAGTTCAGGGGGCTTGAAGTGAAGCGCAATGCAATGAGCAT 4809
Qy      1457 LeuLysPheGlyAerLeuAerProLeuLeuGlyGlnThrGlyLeuProGlnIleProSerGlu 1476
Db      4810 CTGAGATACATACAAAGCCCGGACAGATGCGCTGAGGGGTGACAGGGCCCTCCAAAG 4869
Qy      1477 GlnLysValAlaValValThrValAerGlyCysAerThrGlyValAlaLeuAerPheGly 1496

```

```

Db      4870 GACAAAGTGGCTGTCTAAGCTGTGATGATGTGATGTGGCGCTGCTGCAATTTGGT 4929
Qy      1497 SerValLeuGlyAerPheSerCysAlaIleGlnGlyThrGlnGlyGlySerIleCysSer 1516
Db      4930 GCTGAGATTGGACATACATGCGCGCTGTGTGTGCAAAAGACTCCAAAGAGTCC 4989
Qy      1517 LeuAerLeuThrGlyProLeuLeuLeuGlyValProAerLeuProGlyLeuSerPhePro 1536
Db      4990 CTGAGACTGACGGGCGCTTCTTCTGGGAGGTGTCCCAACTCCCGAGAACATTCGCC 5049
Qy      1537 ValAerMetArgGlnPheValGlyCysMetArgAerLeuGlnValAerSerAerGlnIle 1556
Db      5050 GTATCCCAATAGACATTCAGCTGATGCGGGAGCTGCACATGATGAGCGCGGAGTG 5109
Qy      1557 AserMetAlaAerPheIleAlaAerAerGlyThrValProGlyCysProAlaLysPheAer 1576
Db      5110 GACATGGCGGCTTTGTGTGCAATATGACACATGCGAGCTGCGCAAGCCAAAGTACAC 5169
Qy      1577 ValCysAerSerAerThrCysIleAerGlyGlyThrCysValAerGlnIleThrAerAlaPhe 1596
Db      5170 TTTTGTGACTAGGCCCTGCAAGAACAGTGGCTTCTGCGAGCCCTGGGGGCACTTC 5229
Qy      1597 SerCysGluCysProLeuGlyPheGlyGlyLysSerCysAlaGlnIleMetAlaAerPro 1616
Db      5230 AcTGCAGCTGCGCTGTGGGGCTTCCGGCGCAAGACTGCACTTACTATGGCCCATGCC 5289
Qy      1617 GlnIlePheLeuGlySerSerLeuValAlaThrIle---GlyLeuSerLeuProIleSer 1635
Db      5290 CACCAATTCCTGGCAACGGCACTGAGCTGAATCTTGAAGTGAACATGGCTGTGCTT 5349
Qy      1636 GlnProPhePheLeuSerLeuMetPheAerGlyThrArgGlnAlaAerPheValLeuLeuGln 1655
Db      5350 GTCGATGATGACTGGGGCTGGCAATTCGACACCGGGACAGCGAGGGGTCTGATGCA 5409
Qy      1656 AlaIleThrAerGlyAerSerThrIleThrLeuGlnLeuAerGlyGlyIleValMetLeu 1675
Db      5410 GTGAGCTGGGCGCACAGACAGCTGCTTGCAGCTGATGATCGGGGCTTACGTCTGTG 5469
Qy      1676 SerVal---GlnGlyThrGlyLeuGlnAlaSerSerLeuAerGlyProGlyAerGlnAla 1694
Db      5470 ACGATGACAGAGAGCTCGAGC---GTGCTTCCATCTCTTGTGAACAGGTGACTGTG 5526
Qy      1695 AserAerGlyAerPheIleIleAlaGlnLeuAlaLeu-----GlyAlaSerGlyPro 1712
Db      5527 AGTATGCGCGGTGCAAGTCTGCGCTGAGGTGAGAGGAGAACAGAGTGGCGGGGG 5586
Qy      1713 GlyIleAlaIleLeu-----SerPheAerPheGlyGlnGlnAerGlnGlyAerLeu 1730
Db      5587 GCGCACCAATGCTTATGCTGCTCACTGACTTGAAGCTTTCAGAGACCAATGCGGGTG 5646
Qy      1731 GlyProAerGlyLeuIleLeuIleLeuSerAerIleThrValGlyGlyIle---ProGly 1749
Db      5647 GGAATGAGCTTGCAGAGCTGAAGGTAAAGAGCTCCACGTGGAGGGCTGGCCCCGGGC 5706
Qy      1750 ProAlaGlyValAlaAerGlyPheAerGlyCysLeuGlnGlyValAerValSerAer 1769
Db      5707 AGTGCAGAGAGGCTCCACAGGTCTGTTGCTGCAACAG----- 5748
Qy      1770 ThrProGlnGlyValAerSerLeuAerProSerIleGlyGlySerIleAerValGln 1789
Db      5749 -----CAACCGAGT-----GAATGCGGACT 5769
Qy      1790 GlyCysSerLeuProAerProCysAerSerAerProCysProAlaAerSerIleCysSer 1809
Db      5770 GGTGTGTGTGACCAAGCGCTGTGCTTGGGCGCTTCCCACTGCAGCGCAAGCTGGCGG 5829
Qy      1810 AserAerPheAerSerIleCysSerCysAerProGlyGlyGlyAerAerCysAerThr 1829
Db      5830 GACCTTGCAGACACTTTTCTTGACCTGCGACGAGGTACTAGCGGCCAGCGCTGTG 5889
Qy      1830 AserValCysAerLeuAerProCysGlnIleGlnSerValCysThrAerGlyProSerAla 1849
Db      5890 GATGCTGCTCTGAAACCTGTGCAGAACAGGATATGCTCGGACCTGGACAGAGCC 5949

```

QY 1650 ProHISGLYTYRTHRCYAGLUCYPROBROASNTYRLEUGLYPROTYRCYAGLUTHRARG 1869  
 DB 5950 CCCCATGGCTATACCTGTGATCTGTGTGGGTGATATTCCGGGACCACTGTGACACAGG 6009  
 QY 1870 ILEARGINPROCYPROARGLYTYRTPRGLYHISPROTHRCYAGLYPROCYAASNCYS 1889  
 DB 6010 ATGACACGACGATGCCACGGGGGTGTGGGGAGCCCACTGTGGCCCTGTGACACTGT 6069  
 QY 1890 ASPVALSERLYSGLYPHEASPPROABPCYBAANLYSTHISERGLYGLUCYHISCYLYS 1909  
 DB 6070 GATGTTCACAAAGTTTATTCCTCACTGCAACAAAGAAATGGGCACTGTACCTGCAAG 6129  
 QY 1910 GLUASNHISYTYRATGPROPGLYSERPROTHRCYGLUENUCYASAPCYTYRPROTHR 1929  
 DB 6130 GAGTTCACATACGACGGGGGAGTGACTTGGCTCCCATGTGACTGTACCTGTG 6189  
 QY 1930 GLYSERLEUSERALGVVALCYASAPPROGLUABRGLYGLNCYSPROCYGLYSPROGLYVAL 1949  
 DB 6190 GGCTTCACCTGGCCTCATATGCAACCCCAACAGGGGAGTCCCTGTGCGCCAGAGGCC 6249  
 QY 1950 ILEGLYATGAGLNCYASAPARGCYASAPRANPPOPHAGLVALTHRTHRASGLCYH 1969  
 DB 6250 CTGGCCCGCCAGTGCACAGCTGTGACAGTCTTCGCAAGGTGACAGCCAGCGGCTGC 6309  
 QY 1970 GLUVALASNTYRASPSERCYSPROARGALILEGLUALAGLYLETPTTPROARGTHR 1989  
 DB 6310 CGGGGCTCTATGATGCTGCTCCCTAAGTCCCTAAGATCTGGTGTGTGGTCCCAAGCA 6369  
 QY 1990 ARGHIEGLYLEUPROALALALAPROCYPROLYSGLYSERPHE----- 2004  
 DB 6370 AAGTTTGCGCTCTGCGCACAGTGCCTGTCCCGGGGGGCGCCGGAATGCGGGGTGCA 6429  
 QY 2005 GLYTHRALVALARGHISCYASAPGLNHSARGGLYTRLEUPROPROASNTLEUPHEASN 2024  
 DB 6430 GGTGTGCTGTGCGGCTGTGTATGAGGCCCCAGGTTGGCTGAGCCGACCTTTCAC 6489  
 QY 2025 CYSTHISERILETHRPHESEGLYUENUGLYPHEALAGLUARGLEUINALYARGASGLU 2044  
 DB 6490 TGTACTCTCCCTGCTTCAGAGAGCTCAGTCTGCTGATGAGCTCTAAGACTGAACAG 6549  
 QY 2045 SERGLYUENAPSERGLYARGSERGLNGLNLEUENLEUENLEUENLEUENLEUENLEU 2064  
 DB 6550 ACGGACGTGATGATCATGAGGACCAAGAGCTGTGCTACGCGGTACGCGAGGTGACGTGC 6609  
 QY 2065 HISHTHALAGLYTYRPHIEGLYSEARVALYLSVALALATYRGLNLEUALATHRARGLEU 2084  
 DB 6610 CACACTGACCACTATTTTACGCAAGATGTTTCAAGTCACTGCCCTGCTGTGCCACCTG 6669  
 QY 2085 LEUALAHISGLUSERTHRGHARGLYPHEGLYUENSERALATHRGHINAPVALHISPHE 2104  
 DB 6670 CTGGCTTCGAGAGCATCAGCAGGGCTTCGGGCTGACAGCACAAGAGTCCCACTTC 6729  
 QY 2105 THRGHLEUENLEUENARGVALGLYSEARALALEUENLEUENLEUENLEUENLEUENLEU 2124  
 DB 6730 AATGGAATCTGCTGTGGCGCGCTGTCACTGCTTCCCAAGACAGCGGGAAGCTTGTGG 6789  
 QY 2125 GLUENULE-----GLNGLNTHRGHUGLY-----GLYTHRALATPRLEUENGLNHS 2140  
 DB 6790 GCGGGCGTGGGGGACGGGGCCCTGGGGGCTCCCAAGGACAGCGGGGACTGTGAGGAC 6849  
 QY 2141 TYRGLUALATYRALASERALEUENLEUALAGLNAHWECARGHISHTYRLEUSERPROPHE 2160  
 DB 6850 CTGGAGGATATGACGACCACTCCGCAAGAAATAGGAATCACTCAATCTTAAATCCCATG 6909  
 QY 2161 THRILEVALTHRTPROASNHILEVALILESERVALVALARGLEUABRGLYASAPHE--- 2179  
 DB 6910 GGGGCTGGAGGCTTAATATCACTGTCAAGCATTAACCGCATGAGACCCCAAGTTCTCC 6969  
 QY 2180 ---ALAGLYALAYLEUPROARGTYRGLUALA-----LEUALAGLYGLNGLNPROPRO 2196  
 DB 6970 CGGGGGGCGCGTACCTGCTACCTGCTAACAATAGCAACCTTTCAGAGGCGCAGAGGCGCTGG 7029

QY 2197 ASPLEUGLUTHRTHRVALILEUENPROGLUSERVALPHEARGLUTHRTPROVALVAL 2216  
 DB 7030 GATCTTCACACCCATGTGCTGCTCCCAAGTCCCAAGGCGGACCACTGTAAGTT 7089  
 QY 2217 ARGPROALAGLYPROGLYUWALAGLNGLNPPOGLNGLUENLEUALARGHARGLNAARG 2236  
 DB 7090 CTGCCCAACAGCAGCAGCATGAAATCCACCACTCAAGTGTGTCCCCCAACAGCC 7149  
 QY 2237 HISPROGLYUENUSERGLNGLUWALALASERVALILELETYRATGTHLEUALA 2256  
 DB 7150 CCGCCAGGACCAAGGCTGTGAGTCTCATTAATCTCTCTCTTACCGGACCTTAGGG 7209  
 QY 2257 GLYUENLEUPROHISASNTYRASPPOABRGLYARGSERLEUARGVALPROLYSARGPRO 2276  
 DB 7210 GGACTGCTCCCTCCGAGTCCAGGACAAAGCCGAGTGCAGGCTTCCCTCAAGAACCC 7269  
 QY 2277 ILEILEASHTHRTPROVALVALISERILESERVALHISAPRABRGLUENLEUPROARG 2296  
 DB 7270 GTCATGAACTCCCGGGGTGTGAGCGTGTGTTCACGAGCAACTTCTTAGGGGA 7329  
 QY 2297 ALALEUABRGLYSPROVALTHRTVALGLNPHIEARGLEUENGLUTHRGHARGHLYS 2316  
 DB 7330 ATCTGTGAGTCCCATACAGCTTCAAGATTTCCGTCTCAACAGACAGAAATGGAGCAAG 7389  
 QY 2317 PROILECYVALPHEHTPRASNHISERILEUENVALSERGLYTHRGLYTYRPSERALA 2336  
 DB 7390 GCATCTGTGTGAGTGGAGCCCACTGCGCTGGCGGACAGCATGTGTGTGAGACAGA 7449  
 QY 2337 ARGGLYCYGLUVALVALPHEARGASGLUSERHISVALSERCYAGLNCYBAANHSMEC 2356  
 DB 7450 CGGAGCTGCAGGTGTGTGACAGAAATGGGTCCAGCAGCGGTGTGCTGACCGGACA 7509  
 QY 2357 THSRPHEALVALILEUENLEUENLEUENLEUENLEUENLEUENLEUENLEUENLEU 2374  
 DB 7510 GGAGACCTTTGGGGTCTCTAAGATGAGTCTCCCGTGAAGAGGCGTGAAGGCACTGGAG 7569  
 QY 2375 PROLEULYSTHRTLEUTHRTYRVALALALEUGLYVALTHLEUALALALEUENLEUTHR 2394  
 DB 7570 CTGCTGTGTGTTCACCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7629  
 QY 2395 PHEPHEPHEUTHRTLEUENARGYLEUENARGSERASNDNHSGLYILEARGARGASN 2414  
 DB 7630 GCAAGCATCTGTGAGCTGCCAGCTTCAATCTCAATGTGGGTGAGTTCATGCAAT 7689  
 QY 2415 LEUTHRALALEUENGLYUENALAGLNUENVALPHELEUENGLYILEASNDNALSAP 2434  
 DB 7690 GTGGCAGCGCGCTGGGGGGTGGAGAGCTCTCTCTGTGTGGGATTCACAGAACCCAC 7749  
 QY 2435 LEUPROPHIALACYSTHRTVALILEALILEUENHISPHIEUTHRTLEUCYSTHRTPHE 2454  
 DB 7750 AATCAGCTGTGTGCACTGACAGTGCATCTCTCACTTCTTCTTCCACACACTTC 7809  
 QY 2455 SERTRALALEUENGLUWALALEUENHISLEUTYRARGALALEUTHRGHVALARGAPVAL 2474  
 DB 7810 GCGTGTCTTCTGTGCAAGGGGCTGTCACTTACCGAGTGAAGGTGAGCCACCACTG 7869  
 QY 2475 ASNTHRTGLYPROMEHTARGPHEITYRMELEUENGLYTRPGLYVALPROALPHEILETHR 2494  
 DB 7870 GACGCGGCGCAGCAGCTTCTTACAGGCTCTGGGCTGGGCGCTCTGTGTGCTGCTG 7929  
 QY 2495 GLYUENALVALGLYUENABRPROGLNGLYTYRGLYASNPPOABRPHIECYSTPRLEUSER 2514  
 DB 7930 GCGCTGTGTGTGGCTGTGACCTGTAGGGCTAAGGAACTTCACTTGTGTGTGATCTCA 7989  
 QY 2515 ILETYRASPHTHRTLEUENLEUENLEUENLEUENLEUENLEUENLEUENLEUENLEU 2534  
 DB 7990 GTCCACAGCCCTCATCTGTGAGCTTGTGTGCGCTCTCTGTGTGTGTGTGTGTGTGTGT 8049  
 QY 2535 VALPHELEUTHRTLEUENALALARGALASERCYALA---ALAGHARGLNGLYPHE 2553  
 DB 8050 GGGACATGTTCTCTCTGCTGCGCCACACTCTGCTCCACAGGGCAGAGGGAGGCGC--- 8106  
 QY 2554 GLUUBRGLYSPROVALSERGLYUENGLNPROSERPHEALALEUENLEUENLEUSER 2573

[illegible]

```

Db      9079  GCTGTCGACGAGTGGCTCTTACGATCGCATATCATCTGCGGGAGGACGCGACGACTTTC 9138
QY      2881  ArgProProProProArgInSerLeuGlnGluGlnLeuAsn 2893
Db      9139  CAGCCAGCCAGCCGCTACTCTTCTTAGAGAAACAGCTGAC 9177

RESULT 15
US-09-764-870-569
; Sequence 569: Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT FILING DATE: 2001-01-17
; Prior application date removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 569
; LENGTH: 2332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-569

Alignment Scores:
Pred. NO.:      8,08e-297      Length:      2332
Score:          3958.00      Matches:      776
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      25.46%      Indels:      0
DB:               Gaps:      0

US-09-916-849A-3 (1-2923) x US-09-764-870-569 (1-2332)
QY      254  ThrHisIaIaPheArgValIThrIaIaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273
Db      3    ACCCAAGCTTTCAGAGGTCAAGGCGCAGAGACCAAGCGATGCCCAAGATGCTTGAGCT 62
QY      274  ThrLeuThrIleLeuValIThrAspThrAsnAspHisAspProValIPheGlnGlnGlnIu 293
Db      63  ACACTCAACATCTTGTTACTGACACCAATGACCATGACCTGTGTGTTGAGACACAGAG 122
QY      294  TyrIysGlnSerLeuArgIuAsnLeuGlnValIGlyTyrGlnValIleuThrValArgAla 313
Db      123  TACAAGAGAGAGCTTCAGAGGAACTGGAGGTTGGCTATAGGGGCTCACGTGCAAGGCC 182
QY      314  ThrAspGlyAspAlaProProAlaAsnIleLeuTyrProLeuLeuGlnGlySerGly 333
Db      183  ACGAGATGATATGCCCTCCCAATGCCAATATTTCTGTACGCTGCTGAGAGGGGCTGGG 242
QY      334  GlySerProSerGlnValPheGlnIleAspProArgSerGlyValIleArgThrArgGly 353
Db      243  GGCAGCCCTCTGAAGCTTGTAGATGACCTTCGCTCTGGGGGTGATCCGAACCGTGGC 302
QY      354  ProIaIaAspArgGlnGlnValIGlnSerTyrGlnLeuThrValGlnIaSerAspGlnGly 373
Db      303  CCTGTGATCGGAGAGAGGTGGAATCTTACAGCTGACGGTGAAGGCAAGTACACAGGCT 362
QY      374  ArgAspProGlyProArgSerThrThrAlaIaValPheLeuSerValIGlnAspAspAsn 393
Db      363  CGGAGCCCGGGTCTCGAGATCACCAAGCGCGCTGTTTCTTCTTGTGAGAGATGACAT 422
QY      394  AspAsnAlaProGlnPheSerGlnIuIysArgTyrValValGlnValArgGlnAspValThr 413
Db      423  GATTAATGCCCCCAAGTTTATGTAGAAACGGCTATGTGTCCAGGTGAGAGAGAGTACT 482
QY      414  ProGlyIaIaProValIeuaArgValIThrAlaSerAspArgAspIysGlySerAsnAlaVal 433
Db      483  CCAAGGGGCCCAATGATCTCCAGATCACAGCTTCGATGTGAACACAAGGAGCAATGCCGTG 542
QY      434  ValHisTyrSerIleuMetSerGlyAsnAlaArgGlyGlnPheTyrIleuAspAlaGlnThr 453

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004. Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:52:44, Search time 39 Seconds

(without alignment)  
7207,719 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 15545

Sequence: 1 MRSPTGVPLPTPPPLLL.....ACTVDESSGSEFLFNFLLH 2923

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 segs, 96168682 residues

283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8974.5	57.7	3034	2 T14119	seven-pass transmembrane
2	2491	16.0	2610	2 T20968	hypothetical prote
3	2358	15.2	1364	2 T00250	MEG2 protein - hu
4	1549.5	10.0	5147	1 JUFFTM	cadherin-related t
5	1508.5	9.7	2809	2 T30213	G-cadherin - sea u
6	1421.5	9.1	3097	2 T00021	DN-cadherin - fru
7	1275	8.2	4351	2 T00252	MEG1 protein - ra
8	1092	7.0	4307	2 T20721	hypothetical prote
9	817	5.3	1439	2 T27110	hypothetical prote
10	768.5	4.9	1069	2 T00043	BH-procaderhin p
11	764.5	4.9	1069	2 T00041	BH-procaderhin p
12	729.5	4.7	1467	2 T18411	lactophilin-1, bra
13	729.5	4.7	1472	2 T18413	lactophilin-1, bra
14	728	4.7	1510	2 T17145	CLAB protein - ra
15	728	4.7	1510	2 T17145	CLAB protein - ra
16	727	4.7	1466	2 T17138	CLAB protein - ra
17	727	4.7	1471	2 T17149	CLAB protein - ra
18	727	4.7	1471	2 T17149	CLAB protein - ra
19	727	4.7	1512	2 T18392	lactophilin-3, sp1
20	723.5	4.7	1503	2 T18389	lactophilin-3, sp1
21	723.5	4.7	1571	2 T18395	lactophilin-3, sp1
22	723	4.7	1365	2 S75200	fat protein - syne
23	722	4.6	1283	2 T18394	lactophilin-3, sp1
24	722	4.6	1351	2 T18409	lactophilin-3, sp1
25	719.5	4.6	1463	2 T17159	CL2AC protein - ra
26	719	4.6	1452	2 T17157	CL2AC protein - ra
27	718.5	4.6	1374	2 T18391	lactophilin-3, sp1
28	718.5	4.6	1342	2 T18405	lactophilin-3, sp1
29	718	4.6	1550	2 T14327	alpha-lactocoxin r

30	717.5	4.6	1240	2 T18393	lactophilin-3, sp1
31	717.5	4.6	1308	2 T18408	lactophilin-3, sp1
32	716	4.6	1420	2 T17158	CLAB protein - ra
33	715.5	4.6	1487	2 T14324	alpha-lactocoxin r
34	714	4.6	1231	2 T18390	lactophilin-3, sp1
35	714	4.6	1299	2 T18398	lactophilin-3, sp1
36	712	4.6	1478	2 T17185	CL2BC protein - ra
37	711.5	4.6	1384	2 T18366	lactophilin-2, sp1
38	711.5	4.6	1450	2 T18382	lactophilin-2, sp1
39	711.5	4.6	1467	2 T17160	CLAB protein - ra
40	710.5	4.6	1584	2 T00026	brain-specific ang
41	709	4.6	1397	2 T18377	lactophilin-2 (sp1)
42	709	4.6	1463	2 T18386	lactophilin-2 (sp1)
43	708.5	4.6	1435	2 T46611	CL2BB protein - ra
44	708	4.6	1341	2 T18301	lactophilin-2, sp1
45	708	4.6	1341	2 T17200	CL2BC protein - ra

#### ALIGNMENTS

RESULT 1					
T14119					
seven-pass transmembrane receptor protein precursor - mouse					
C:Species: Mus musculus (house mouse)					
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999					
C:Accession: T14119					
R:Hadjantonakis, A.K.; Formstone, C.J.; Little, P.F.R.					
submitted to the EMBL Data Library, October 1997					
A:Description: The Celser family of novel evolutionarily conserved seven-pass transmembra					
A:Reference number: Z17881					
A:Accession: T14119					
A:Status: preliminary; translated from GB/EMBL/DBJ					
A:Molecule type: mRNA					
A:Residues: 1-3034 <HMD>					
A:Cross-references: EMBL:AF031572; NID:g3800735; PID:g3800736; PIDN:AAC68836.1					
C:Keywords: transmembrane protein					
F:1-26/Domain: signal sequence #status predicted <SIG>					
F:27-3034/Product: seven-pass transmembrane receptor protein #status predicted <MAT>					
Query Match					
Beet Local Similarity 57.7%; Score 8974.5; DB 2; Length 3034;					
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;					
QY	1	MRSPTGVPLPTPPPLLL	1	PPPLGDDGPCRSLSRGGS	48
DB	124	LRRSARGAELRSP	1	AVRSVPLGDLAL	172
QY	49	GACAPM	1	PSSASNLMLYTSRCDAGTETLGHVPHHDGLRWCPSSEA	99
DB	173	CSCPVNAGTCGRPRPICRPGSABLRVLCALGAAGA	1	VWV	213
QY	100	HIPLPAPBGCPCWCRLLIGIGHLSPQKLTLPESHPLKAPRLRCOSCKLAQAPGLRAG	1	159	
DB	214	ELVYQTSSTPESPESV	1	SPSL	243
QY	160	ERSPESLGRKRKNNTAPQFQPSYQATVENQAGCPVASTRAIDDDGAGLE	1	217	
DB	244	VVRNS	1	RRGTGSSSTSPQPLPSYQVSPENEPAGTAIVBLRHADPBGAGLS	296
QY	218	YTMALPDRSNOPFSLDPVGAATTAETLREKTSVTFVTAQDQHGMPRRSLATLTI	1	277	
DB	297	YQWELPDRSNGVFLDAATGATTAASLDREKTDHVLKVASVDHGSPPRSATVLTIV	1	356	
QY	278	LVTDTNHDVPYBQEQEYKESLRNLENGYEVLTTRATDGDAPPNANITRLLEGSGGPS	1	337	
DB	357	TVSDTNDSPVFBQSEYRERIRENLENGYEVLTTRATDGDAPSNANMYRLLEGSGG	1	413	
QY	338	EVFEIDPSSGVIIRGPVDRREESYQLTVEASQDGRPGRRSTAAVFLSVEDNDNDAP	1	397	

Db 414 -VEIDARSGVTRAVVDBEAAIYQLVEANDOGNPGPLASATVHI VVEDENDTP 472  
 QY 398 QFSEKRYVQVREDVTPGAEVLRTASDRDKGSNAVHVSIMGNARGOFPYLDACGTALD 457  
 Db 473 QFSEKRYVQVREDVAVNTAVLRTAQATDRDQGNAAIHVSITSGNKGQFYLHLSGSLD 532  
 QY 458 VSPFLDYETTKETLAVRADGGRPLSNVSGVTVYQVLDINDNAPFVSTPQATVLES 517  
 Db 533 VNPDLFEALREXYTLIKADQDGRPLINSGLVSVQVLDVNDNAPFVSSPQOALVLEN 592  
 QY 518 VPLGYLVHQAIDADAGNARLERYLAGVH-----DFFPTINGTG 560  
 Db 593 VPLGSHVLIHQADADAGNARLQYRLVDTASTIVGSSVDSKNPASPDPFQOLHNSSG 652  
 QY 561 MISVAALDEEVDYFSPGYEABDHGTALTAASVSVTVLDVNDNPFPTQPEYTVRLN 620  
 Db 653 WITVCALDEEVEHVSFGYEAVDHGSFAMSSASVSTVLDVNDNPFMTQGVYELRLN 712  
 QY 621 EDAAVGTSVTVANVDRDASVTITQITSGTNRBSITSGSGGLVSLALPDYCLEHQ 680  
 Db 713 EDAAVGSVTVTLARRDANSVTITQITGNTNRRLSSQSGGLITLALPDYQOERQ 772  
 QY 681 YVLAATASDGTRODTAQIVNVNTDANTHRPVFQSSHVTVVNNEDRPAGTVVLISATDE 740  
 Db 773 YVLAATASDGTSHTAQVFNVTIDANTHRPVFQSSHVTVSVDREPVGISATISATDE 832  
 QY 741 TGENARITYMEDSIPOFRIDADTGAVTTOQABLDEDOVSYTLATARDNGIPQKSDTTY 800  
 Db 833 TGENARITYVLEDPVPOFRIDPTGTITVTELEDEDOAAVYTLATARDNGIPQKSDTTS 892  
 QY 801 LETLVNDVNDNAQOFLRDSVQSGVYEDVPEFTSVLQISATDRSGANGRYFTYFQCGDNG 860  
 Db 893 LETLVNDADNARFLRDFYQSGVFEDAPSTSVLQVSAIDRDSGNGRLLYTFQCGDNG 952  
 QY 861 DGDPIVESTSGIVRTLRDLRENNVAQVLAAYVADKGM-PARTPMEVTVTVLDVNDNP 919  
 Db 953 DGDPIEPTISGVARTQRDLRENNVAAYVNLALAVDRSGNPLSASVIGVSTVDINDNP 1012  
 QY 920 VPEODEPVEBENSPIGLAAVAVATADPDDEGTNAQIMYOIVEGNIPEVFOQDIFSGEILT 979  
 Db 1013 VPEKDELELFEVBENSFVGSVVAARIRANDPDEGPAQIYQIVEGANVEVQDILLSGDLR 1072  
 QY 980 ALVDLDEDEPEVTVYQATSAPLVSRATVHVLLBRNDNPVLGNPEILLFNNYNNRBS 1039  
 Db 1073 ALVELEDEVRDYMVLVQATSAPLVSRATVHRLQNDNPPELPOLLFNNYNNRBS 1132  
 QY 1040 SPPGCAIGRPADHPDISLSTYSPERGNELSVLNLASTGEKLGRALDNNRPLAINS 1099  
 Db 1133 SPPSGVIGRIPADHPDISLSTYFLOGNELSLILLDPATGELQSLRDLNRRPLAINS 1192  
 QY 1100 VLVSQVHSTYAOACALRVITITDEMLTHSITLLEDMSPERFLSPLIGLFIQAVATLAT 1159  
 Db 1193 VVSVDGHSYVALCTLRVTIITDDMLTNSITVRELENNSGEKLSPILLSTFVSGVATVLS 1252  
 QY 1160 PRDHVVFNORPDACGHILNVSLSVQPPGCGGPPPLSEDOERLXNRSILTAI 1219  
 Db 1253 TKODIFENQONDV-SSNLTANTFSAIIPGGRG--RFPSEBDQEQIYLRNLTIT 1309  
 QY 1220 SAORVLPFDNICLREPCENYMCVSLRFDSSAPFJASSSVLFRPIHPGRLCRCPG 1279  
 Db 1310 SAORVLPFDNICLREPCENYMCVSLRFDSSAPFJASSSVLFRPIHPGRLCRCPG 1369  
 QY 1280 FTGQYCEVEYDLCSRPCGPHGRCSRREGYTCLCRDYGTEHCEVARSGRCTPGVCKN 1339  
 Db 1370 FTGQYCEVEYDLCSRPCGPHGRCSRREGYTCECFDFTGEHCQVNVRSGRCAISGCKN 1429  
 QY 1340 GGTGVNLVSGPKDCSPGPKPYCQVTRSPFPAHSFIFRGLRORFHTLALSPATKE 1399  
 Db 1430 GGTGVNLVSGPKDCSPGPKPYCQVTRSPFPAHSFIFRGLRORFHTLALSPATKE 1489  
 QY 1400 RDGILLNNGRNEKHDVLEVALQEQVOLTFSAGSSTTVSPVPGVSDGQWHTVOLKY 1459  
 Db 1490 RNALLNNGRNEKHDVLEVALQEQVOLTFSAGSSTTVTPQVPGVSDGQWHTVOLKY 1549

QY 1460 YNKPILLGQITGPEQGBEOKVAVTVYDGDYVALRFGSVLGNYSCAAQGTQSGSKSIDL 1519  
 Db 1550 YNKPINIGHLPHGSGSKVAVTVYDDCAVAVHFGSVYGNYSCAAQGTQSGSKSIDL 1609  
 QY 1520 TGPILLGGVPLDPSEFPVRMQPVGCNRNLOVDSHIMADPIANGVIRPGCAPKQVCD 1579  
 Db 1610 TGPILLGGVPLDPSEFPVRMQPVGCNRNLOVDSHIMADPIANGVIRPGCAPKQVCD 1669  
 QY 1580 SNTCANGTQVQOMAFSCCEPLGFGKSCAOMENPQFLGSSIVAHGSLPISQPMY 1639  
 Db 1670 GTSQNGGTCTNKRMTYLCCEPLRGFGKSCAOMENPQFLGSSIVAHGSLPISQPMY 1729  
 QY 1640 LSLMFRTOADGVLLQALTRGSRITTLQLRGHNVLSEVGTGLQASSLRLBERANDGDW 1699  
 Db 1730 LSLMFRTOADGVLLQALTRGSRITTLQLRGHNVLSEVGTGLQASSLRLBERANDGDW 1789  
 QY 1700 HHAQALALASGPRG-----HLSFDYQQAABGRLGLHLSNTTVGGIPEPAGV 1754  
 Db 1790 HHLLEL-RSAXBGKDIKYLAVMTLIDYGMDSITVQIGNQLPGLKRTTVIGGVTEKVS 1848  
 QY 1755 ARGFRCLOGVVSDTPGVSILDPHGESINVEQCSLPPPCDNPSPANSYCSNDMS 1814  
 Db 1849 RHGFRCLOGVVSDTPGVSILDPHGESINVEQCSLPPPCDNPSPANSYCSNDMS 1908  
 QY 1815 YSCSCDPGYGDNCTNVCNLPCEHQSVCTKPSAPHGTYCECPNYLAGPYCEIRIDPC 1874  
 Db 1909 YSCICDRGYFKKCYDACILNPKCHVAACVSPNTPRGSCGCGHGYQYCEKNTVADPC 1968  
 QY 1875 PRGMWGHPTCPGNCNDVSKGPDCKNTSGRCHCKENHRRPGSPTCLCCQYPGISLR 1934  
 Db 1969 PRGMWGHPTCPGNCNDVSKGPDCKNTSGRCHCKENHRRPGSPTCLCCQYPGISLR 2028  
 QY 1935 VCDPEBDGQCPCKPGVIGRQCDRCNPFABVTTNGCEVNYDSCPRALTEAGIWMPTFRGLP 1994  
 Db 2028 ACMDMTGCAKCPGVIGRQCDRCNPFABVTTNGCEVNYDSCPRALTEAGIWMPTFRGLP 2088  
 QY 1995 AAAPCPKSPGTAVHCDHGRGMLPNTFNCTSTPFSBLKGFABRLQNNESGLDSGRSQO 2054  
 Db 2089 AAAPCPKSPGTAVHCDHGRGMLPNTFNCTSTPFSBLKGFABRLQNNESGLDSGRSQO 2148  
 QY 2055 LALLRNMTOTFAGVFGSDYKVAVOLATRLAHESRTOFGFSLATODVHFTENLRVSA 2114  
 Db 2149 LALLRNMTOTFAGVFGSDYKVAVOLATRLAHESRTOFGFSLATODVHFTENLRVSA 2208  
 QY 2115 LLDTANKRMWELIQQTEGTAWLQHYEAYASALAQNNRHTYLSPTITVPIVLSVRL 2174  
 Db 2209 LLDTANKRMWELIQQTEGTAWLQHYEAYASALAQNNRHTYLSPTITVPIVLSVRL 2268  
 QY 2175 DKGNFAGAKLPRYBALRGEOPDLETTYLPBSVF-----RETPPVYR-----PAGP 2221  
 Db 2269 DKGNFAGAKLPRYBALRGEOPDLETTYLPBSVF-----RETPPVYR-----PAGP 2328  
 QY 2222 GEAQPEBELARORHPELSQGEAAVASYIYRTLAGLLPHNYDPKRSIRVPRKPIINTP 2281  
 Db 2329 GEAQPEBELARORHPELSQGEAAVASYIYRTLAGLLPHNYDPKRSIRVPRKPIINTP 2388  
 QY 2282 VVSISVHDEBELPRALDKPVTVQFRLLETERKPICVFNNHSLVSGTGSMSARCEV 2341  
 Db 2389 VVSISVHDEBELPRALDKPVTVQFRLLETERKPICVFNNHSLVSGTGSMSARCEV 2448  
 QY 2342 VFRNESHVSQCCNHTSPAVLMDVRRRNGELLPKLTLYVALGYTLLALLLTFPFLYL 2401  
 Db 2449 VFRNESHVSQCCNHTSPAVLMDVRRRNGELLPKLTLYVALGYTLLALLLTFPFLYL 2508  
 QY 2402 RILRSNOGIRBNLRAALGAQVFLGIGNQADLPACTVIALIHLHFLYLCFSSVALLEA 2461  
 Db 2509 RILRSNOGIRBNLRAALGAQVFLGIGNQADLPACTVIALIHLHFLYLCFSSVALLEA 2568  
 QY 2462 LHLRYALTEVRDVTNGPMRFYTMGLGWPAFTYGLAVGIDPBGYANPDFCWLSTYDTLW 2521  
 Db 2569 LHLRYALTEVRDVTNGPMRFYTMGLGWPAFTYGLAVGIDPBGYANPDFCWLSTYDTLW 2628

QY 2522 SPAGPVAFAVSVVFLYIIAARASCAAQOGFEKGPVSGLOPSPAVLIIILSLATWIIALL 2581  
 DB 2629 SPAGPVGTYIIITVTVLIVLFAKVSQCKHRYERKGVSVMLRTAFILIIIVATWIIIGLL 2688  
 QY 2582 SVASDILLFHYLPATCNICIGPFILSYVVLKSEVRAKAL-LACSRKPSDPAIITTKSR- 2639  
 DB 2689 AVNSDILSRITLPAASCLQGITVLLPHCVANREVRHMLAVALAGKQLQIDDSATTATLV 2748  
 QY 2640 LTRSYNCPSPYADG--RLYOPYDSDAGSLHSTRSGKSPSYIPLLRBSALNPGQGP 2697  
 DB 2749 LTRSLNCNNVTYSGPMLRTALGSESTASLDSTRDGVQ-----KLSTSSGPANG 2798  
 QY 2698 GLGDPSGLPL-EGDQDQHPDPTSDSLIEDQSGSYASTSSDSEEEBEEBEEBAP 2756  
 DB 2799 NHEBPTSTPLPNSKKAHGDSDSDRLSL-DHSSSYASTSHTSDSDGGEAEADK----- 2853  
 QY 2757 GBOGMDSLGPGAEPLPLSTPK-DGPGPGKAPWBGD--FGTTAR----- 2800  
 DB 2854 ----WNPAGGPA-----HSTPADLALANHVPAWGPDESLAGSDSELDTEPHLKVETV 2903  
 QY 2801 -----SSGNAPBERLRNENGDAISREGSLGPLPGSSAQP--HKGIKKKCL--PTIS 2848  
 DB 2904 SVELHRAQAGNHCGRPSDPBESGLAK-----FVAVLSSQPGQRKGIKKKVTYPPPLP 2958  
 QY 2849 EK--SSLPLPLEQCTGSSRGS-SASBGRGCPP-----RPPRQSLQEQIANGVMDI 2898  
 DB 2959 EPLKSRRLREKADCEOSPTSSRTSLGSDGVHATDVTITKTPRPRERGRHILNGV--- 3015  
 QY 2899 AMSIXAGTVDEDSGSGSE 2915  
 DB 3016 AMNVRTGSAQANGDSSE 3032

## RESULT 2

T20968  
 Hypothetical protein f15B9.7 - Caenorhabditis elegans  
 C.Species: Caenorhabditis elegans  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C.Accession: T20968; T26278  
 R.Pericy, C.  
 Submitted to the EMBL Data Library, August 1996  
 A.Reference number: Z19351  
 A.Accession: T20968  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-2610 <M1>  
 A.Cross-references: EMBL:Z78013; PIDN:CAB01427.1; GSPDB:GN00023; CESP:F15B9.7  
 A.Experimental source: clone F15B9  
 R.Baynes, C.  
 Submitted to the EMBL Data Library, August 1996  
 A.Reference number: Z20187  
 A.Accession: T26278  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-2610 <M1>  
 A.Cross-references: EMBL:Z78018; PIDN:CAB01449.1; GSPDB:GN00023; CESP:F15B9.7  
 A.Experimental source: clone W0764  
 C.Genetics:  
 A.Gene: CESP:F15B9.7  
 A.Map position: 5  
 A.Introns: 169/2; 217/3; 239/3; 284/3; 515/2; 550/3; 738/3; 810/1; 838/3; 890/2; 977/1;

Query Match 16.0%; Score 2491; DB 2; Length 2610;  
 Best Local Similarity 27.4%; Pred. No. 8.2e-119;  
 Matches 768; Conservative 438; Mismatches 1205; Indels 394; Gaps 89;  
 QY 51 CAPMGTLCCSSASNLMLYSR--CRDAGTBLCHLVPHHDGLRWCPESBAHPLPAPAE 108  
 DB 39 CRPCA--VSSNSVIVLPASRRCPLHGPQPII-HMPDLSDMLA-----CPVPGJPD 86  
 QY 109 GCPWSCRLLGIGLHSPGKGLTPEEHPC-LKAP-----RLRQSCKLQAQAPGL 156  
 DB 87 SVHSSQISL-----LEGRG-LILTKERICFPQGPIDPHDYCDGCLYNSKRIHGISIAS 140

QY 157 RAGERSPEESLGGRRKRWNTAPQFPSPYQATVENOPAGTPVASLRAIDDEGEAGRL 216  
 DB 141 KKKLETRTKWRARRRNDANAVHGOEKYVLEBEDPIETIIISAVK---SHASQPL 197  
 QY 217 EYTDALPDSRNOFPSLDPVTGAVTTABELDRFTKSTHFVRYTAODHGMRRSALATLT 276  
 DB 198 YSMAVAPQDSRQNFPTLDWSGEIRLAKSMREVLIDKIIILYATVERVDPITISATTV 257  
 QY 277 ILVTDNDHPVEEOGKESRENLVEYEVLTATRDGDGAPRANILYRLDESGGSP 336  
 DB 258 VAVLDVQNSPLFEKOSTFGBIREDAPIGTIVLSPADLDGKENGELYSIGENG--- 314  
 QY 337 SEVEFIDPRSGVTRFGPDEEVESTOLTYEASDQGRDPGRSTAAVFLSVEEDNDNA 396  
 DB 315 KMLAINAKSGVIGTAAPLDRETLILIDVIAKGP--KRESTANVEITVVDNDNA 372  
 QY 397 POPSEKRYVOVREVTGAPVLTATASDRDKSAAVHYISGSGNANGQVTLAQGTAL 456  
 DB 373 PVFASDSYVNTILENITPAVATATKATDEDFGTNGKHYSWASSSGIGLTIIDYSTGEV 432  
 QY 457 DVVSPLDYETTEKYLTVRAQDGRPLSNVSGLVTVGLDINDNAPIVSTPPQATYLE 516  
 DB 433 TIKERIDAKNS-PITAVIRAKDGAQALSTVPL-TINVIDINDAPFLIAQKMITEE 490  
 QY 517 SVPLGLVLYVQALDADADNARLEYRLAGVGHDPFTINGGWIISVAABIDREEDVDF 576  
 DB 491 NVALIGEEVGRVYALDEDSGPGNIIKYSMEG---SDFIIDDSGLIKTKLDRRTARY 547  
 QY 577 SPGEARDHGTPALTASASVSTVLDVNDNPTFPQPYTVRLNDAAVGTSVTVSAVD 636  
 DB 548 SLKVTARDMGTSLNTSTTIAVVLIDINDNAPTEPKKEYNVTISEMRGSOIITLKAVD 607  
 QY 637 RDAHSVITYQITSGNTRRFSITSGGGGLVSLPLVYKLERO---VLAATYASD-GTR 692  
 DB 608 NDEDKITTYIRERAD-KEVFSIILDGQAL---LSVGELRKQDHKRVATISATDQGL 663  
 QY 693 OPTAQIVNVTDANTHREVPFQSHYTVVNDPRAGTVVLIS----- 735  
 DB 664 QQRCVANVFIDVNS-APVFNHPPSVKIPHSPIGYVITLKVAFSGVFGYGLKHFFV 722  
 QY 736 -----ANDEDTGERARTTYEMEDSIPOFRIDADTGAVTTOALDYEDOV 779  
 DB 723 SSNLRKKRIIPNSAMWMAEDHGRGNARIVYGI-DSSQFRIDPDSGDSVSDLDREBRA 781  
 QY 760 SYTLAITARNDGIPKSDTYTLEILVNDVNDNAPDLASDQSGSYVEDVPPTSVLQISA 839  
 DB 762 TFSVAVTASDASPPPLNTSTQIEVLDDINDNSPQFTSSSYAATISSEDIPIVGTSLQVSA 841  
 QY 840 TDRSGLNGRVYTFQGGDDGDG--FVSESTSGIVRTLRLIDRENVAAQVYLRAVAVKG 897  
 DB 842 IDADIGPAGIVDYFLNBSGSSPSIQLPFLDRTSGLTARVSSKLDREGQFVTVLPFARBG 901  
 QY 898 MPRAATPMEVTVTVLVNDNPPVFEQDFVDVFBENSITGLAVARVTTATDDEGNTAQIM 957  
 DB 902 TFSLSABEITLTLSDVNDNAPTEQLSYDYLIENSVGSITGVITVARDDEGNADIS 961  
 QY 958 YQIVEGNIPEVFOIDIFSGE-----LTAIVLDYEDRP-EVYLVVQATSAPVSRATYHV 1011  
 DB 962 FRIBGADAKLFDIBEDAKQNGVRIILTRAFTDEAKANKFFELQASSGQSLSTVPVRI 1021  
 QY 1012 RLIDNDNPVLGNFEILFNNTVNRSSSPFGAIGRPVADPDISGLTYSFPERGELS 1071  
 DB 1022 HVSVDVNDKPKLPKQFVILNRFVNDVQMAR-----QIGFIPARDPDONATLRYFLB---END 1074  
 QY 1072 LVLLNASTGELKLSALDNRPLRLAKSVVSDGVHSTTAQACARVITITBMLTHSTTL 1131  
 DB 1075 LIAEKYTKLIVQEWGRN--MDVSFKTCVSDANRECSCTRFTHLVBEWMSSESTPL 1132  
 QY 1132 RLEDMSPERFLSPILGLEIQAATLAT-PPDHVVVFVVGQDTAPGGHILNVSLSVQCP 1190  
 DB 1133 SLARTTVDDFMDPLVFGFRDAMSTLSNWKPSDHIHVGKHLD-DVIYINIAITLIDGR 1190

```

Qy 1191 PGPGGPPFPESDQLRLYLNRSLTAISAORVLPDDNLCLEPCENMRCVSLRFD 1250
Db 1191 VVGWRAIIELVKSIKK--LEKWTLLQVEYIR-----DESCANEPGSHAKCQKQKFV 1242
Qy 1251 SSAPFIASSSVLPRIHPVGLRCRCPGPT-----GDYCETEDLCISRPCGHEGRS 1305
Db 1243 GEMKAMHTDPIATLTATVTMTFVCEPCSGFTSGANBD-CDTRIDECYRRCSSNNSVCVA 1301
Qy 1306 REGGYTCLRDYGEHCEVASASGRCTPGVCKNGTGVNLLVGGFKC-DGP-SGDFEKP 1363
Db 1302 FENTYQCECKEKGWIRGICEISVHALTCVPKCSNDSIC-ELDKQMKCRCKHGEPTDE 1360
Qy 1364 YCQVTTSPFASPIFERGLRQRFHTLASFATKEDGELLVNGRPNKHDVVALLEVIQ 1423
Db 1361 RCRLRSVSPDEGELANVLDLPRTQMTKPFVSTIANHGVLVFTG--DKSDPFVEGVVD 1418
Qy 1424 EQVQVLTSSAG--ESTTTSVPVPGVSDGQWHTTQKTKYKPLLGQGLDQGBEGRVAV 1481
Db 1419 RVLKQVPSLGGEEKIDAKMENDVENRINDGWHVVALREYSNKQI-----T 1462
Qy 1482 VTVNGCCTGVALRFSGVILANYSACAGTGGGSKS-----LDLTGPLLGGVPLD 1531
Db 1463 MSLLDDCETNBSLLANT---SPNCAIRAKNLKCEPPTVPCRYLIDISNGLPLGGRPBT 1519
Qy 1532 PESFPVRQDPVGCNRLQVDSRHIDMAF--IANNQTV-PGCPAKXNCDSN--TCANG 1586
Db 1520 SKQJE---KAFGCISLSDVDKEDVDFTSIKEMHKVQVHGCGHGRDQFSTSDQCSAT 1576
Qy 1587 GTCYNQNDASCEPPLRFGGSCAQEBANRPHFGSSLVAMHGLSL----PIGQPTLSL 1642
Db 1577 SKCNRRNGRISCSPQSVHSTGECVAGLTQDLRGHSLFEEBSFVLQPSQVSVPPFVSF 1636
Qy 1643 MFTROADGVLLQAITRGSRITTLQLREGHVMSLV-BGT---GLQASLRLPERANDGD 1698
Db 1637 BFRSRAD---MOVF-----ALEFTORSVHNLEVDQDTLKNVIGDBEVELPABEVTSKH 1688
Qy 1699 WHEAQLGASGGFHAHLSFDYGOQRAEGNGLGRHLGLHSNITVGGIRPGAGVARGF 1758
Db 1689 WNNVVIIFRAD---SVATISINGIYAAEAKAS--ISDMNIESLYFGIADGT--GHSRF 1739
Qy 1759 RGCLQGRVSDTPRGVNSLDPSHGESINVEQ-----GCSLDPDGD-SNCPANSYCSND 1811
Db 1740 EGCIRNLVLD-----GRSISVKKKKGKTRAGCAVPPNCSVDISICPABS----- 1781
Qy 1812 WDSYSCDDEPYGDNCTNVCDL-NPEHOSVCTRKPSAPHGTYCECPMYLAGPYCETRI 1870
Db 1782 -----DTCLPVSVANVCSSTGCVSNNTA--GYCICIPAGKTGNCCDLBA 1825
Qy 1871 -DQCPRGWNG-HPTCPGCNCDVSKGPDPCNKTSGECHKENHYRPPGSPFTCLDCYP 1928
Db 1826 PKQMCPSGMMGTFFRCRRCSCAQKTKDYEAQCCKTKGACQCKKSHSTING--CYKCEC-G 1882
Qy 1929 TGSISRYCDPEPDGQCPCKPGVITGRQCRC---DNPPAEVTNNGEYVYDGCPRALIEGIM 1985
Db 1883 FGAOSTEBS-ADGCKCKNGAVRRCRCSRFHQDLDSKTLK-CRVSIGKPSIEFSIQ 1940
Qy 1986 WPTTRFGLPAAPCPKSGFTGAVNHCDEHRCMLPPLFNCTISITFSELKGAERLQHNES 2045
Db 1941 WPAQKGSIVYQSCPVESGLATKCIETGMSVNVNMANCRRPREYSIMVKEFELR----- 1996
Qy 2046 GLDSGRSQOALLIRNATQHTAGYFGSDVAVAYQALATRLLAHSTORGFGLSATQVHPT 2105
Db 1997 -----PSKILTYANNTNTSSIRGRNOQIAAALSLVDYEOSMPEMKRAHIKDMKFT 2050
Qy 2106 ENLLR-VGSALLDPAANKHMEILIOOTBGTAMLLQHYEAVASALAONMHTYLSPTIYT 2164
Db 2051 EKLIESIGRWSEOPADEYSTLISK-----LNNYAEVYAEIHEAV--NFSLSPFYVAN 2100
Qy 2165 PNIVISVRLDKGPAKALPRYEAALRGEOPDLETTVILPEVSFRETTPVPRVADGGEA 2224
Db 2101 DHIVFASDKLDPGNI---LPKFNMF-----VDLRPTGF-----PRVRAIVAGTT 2141
Qy 2225 QEPBELARQRHREBELSGGEAVAVIYRTLAGLLPHNYDPRKSLAVPKRPIINTVEVS 2284

```

```

Db 2142 Q-----VYYSIVFY-----RCRCRCENPMIA 2162
Qy 2285 ISVHDEBELLPRALDKPYTVQORLLETERTPICVPMHSHILVSGTGMARGCVVR 2344
Db 2163 IVANTSD-----PVIVPEIEEDDGMKYPECCRPDEK-----SGWTYAGALLIGU 2208
Qy 2345 NESHVSCQCNHMTSFAYLADVSRRENGEILPKTLTYVVL-GVTLAALLTFEFTLLRI 2403
Db 2209 NLTHACENRIGVLFMFVNDQSSIVRAVQMDNMTSPALAGVALPLCTSLITLSRRS 2268
Qy 2404 LRSNQGIRRNITLALGALQVFL--LGINQADLPACTVIALILHFLYLCFSSALLER 2461
Db 2269 LKTHSVRIGFILLFPAINIINTLFVHTALNOA-----YCEVRNAMSFTSSAPFAMFLYG 2324
Qy 2462 LHLVRLTEVRDVTNGPMEFYMLGKGVPAFTGLAVG-----LDPEGNDPDCMLSI 2515
Db 2325 LYIYRRLAD-----GSSPSLTSLSLVGIVPCLLSFTTFVTDQCSLSP 2369
Qy 2516 YDTLWSFAPVPAFAVMSVPLYTLAARASCAQRGFEKGFVSGIAPVALLLSAT 2575
Db 2370 HMLFPCILLPLIGLFIILSFYAAASVLVSLHKXDVVFVAKNVKRAVQHFILITFTG 2429
Qy 2576 WLIALISVNSDTLLFHYLPATNCICQPTFLSYVVLSTKVRKALKACSRKPS---PD 2631
Db 2430 MTLTGLFANQLPLMEIMEISQSI---YLIALVIFLWCVCDDITTKADSDNPSMULDQ 2486
Qy 2632 PALTTSTLSSYNCPSPYADG-----LYQYGDASGLSTSRSGSCP----- 2677
Db 2487 KSVMAESTNADP-QCASPLSRPHQHHRVPMDSBWPVDPVPSNHTH-TSINEPDTIRLL 2544
Qy 2678 -----SYIPFLREBSALNPGQPPGLADPGSL-FLBGQDQDHD 2715
Db 2545 LPQNRDVINILSPDQILNKGVGHVYRNMGSLPRLRSQDEAD 2589

RESULT 3
T00250
MEGF2 protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C/Accession: T00250
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomes 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00250
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1364 <NAK>
A:Cross-references: EMBL:AB011536; NID:93449297; PIDN:BA32464.1; PID:93449298
A:Experimental source: brain; clone HG1044
C:Genetics:
A:Gene: MEGF2
A:Map position: 3p21.2-p24.1
C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homo
F:1-28/Domain: EGF homology (fragment) <EGF>
F:32-66/Domain: EGF homology <EGF1>
F:124-169/Domain: laminin-type EGF-like homology <LGB>

Query Match 15.2%; Score 2358; DB 2; Length 1364;
Best local similarity 42.3%; Pred. No. 2e-112;
Matches 495; Conservative 180; Mismatches 372; Indels 124; Gaps 25;

Qy 1801 PCPANSYCSNDMSYSCDPPGYGDNCTNVCDLNCEHOSVCTRKPSAPHGTYCECPN 1860
Db 1 PCPNAHDCRDLMTGTSCTQPGITGPGCYDACLINPCQNGSSRHLPGAPHGTTCCVGG 60
Qy 1861 YLAPYCETRIDQCPRGWNGHPTCPGCNCDVSKGPDPCNKTSGECHKENHYRPPGSPFT 1920
Db 61 YFGHHCENRMDQCPRGWNGSPFCGFCNCDVHKGFDPNCKNTKGQCKEFTYRPRGSDS 120.
Qy 1921 CLLDCYCPGSLRVCDPDDGQCPCKPGVITGRQCRCNDNPPAEVTNNGEYVYDSCPRAL 1980

```

```

Db      121 CLECDYCPVSTSRSCAPHSGQCPRCGALGRQCNCSDFPFAVYASGCVLADACPKSI 180
Qy      1981 EAGIWPRTFGLPAAPACPKGSF-----GTAVRHCDEHGMTPPMLFNTSTTFELNG 2035
Db      181 RSGVWMPQIKETGLATVPCRGALGRLGAGAAVALCDEAGMKEPDLFNTSTPAFELSL 240
Qy      2036 PAERLORNESGDLGSGSOQLALLRNATOTAGYFGSDVAVAYOLATRLAHSESTORFEG 2095
Db      241 LLDGLINKALDITMEAKQAQRLREVTGHTDHFSDQDVAVTARLALHILAFESHQGRG 300
Qy      2096 LSAOTGVHTFENTLRRGSAALLDPANKHMLI-QQTEG---GRAMLIQTEAVASALAQN 2151
Db      301 LVTQAQHFENMLMAGSALLABETGDLWALQGRAPGSGAGSVRHEKTAATLARN 360
Qy      2152 MRHTYSPPTIVTIVTIVSVRLDKGNF--AGAKLPRYEA--LRGEOPDLFTVTLIPES 2207
Db      361 MELTYLNPMLVTPNIMLSIDRKEHPSFPGARRYRHSNLRGQDAMDPTHVLIPSQ 420
Qy      2208 VFERTEPPVVRPAGPGEAEPBELARRORHPELSQENAVSVIITLGLLPHNYDPK 2267
Db      421 SPAPSESEVLPSTSSISNTSTSSVPPAPPEPEPGISIIILVYRTIGSLPAQOAR 480
Qy      2268 RSLRVRKRPRIINTPVVISVHDBELLPRALDXPVTVQFRLTEERTKTCYFNNHSLI 2327
Db      481 RGARLPQNPWNSPVVSAVFRGNFLRGILBESPISELEFLIQTANKSKAICVQMPDPI 540
Qy      2328 VSGTGVMSARGCEVVFRENESHVSQCQNHMTSPAVLMDVSRRE--NGEILPLKTLTYVALG 2385
Db      541 AEHGHWATARDCLVTRNGSHACRCGRITTFGLMDASREELBGLLAVFTHVVA 600
Qy      2386 VTLAALLTFPFLTLRLINSNOHGIRRNLTALGLAQLVFLGINQADLPACTYATLI 2445
Db      601 VSAVALVLTALIIISRLSKSNVRGIHANVAALGVAELIFLGIHRTNQVCTAVAIL 660
Qy      2446 LHELVICTSMALLELHLRYALTVEADVNTGEMRPTMYLGNVPAFTGLANGDPEGY 2505
Db      661 LHFVFLSTAVMLVVOGILHTRMQLVEPNVDGRMRFTHALGWSVPVTLGLAAGDPEGY 720
Qy      2506 GNDPFCMLSTYDTLINSFAPVPAFAVMSVFLYTLAARASCA--AORQGEFKGPVSGLO 2564
Db      721 GNDPFCMISVHEPLINSFAPVPAVIVMNGTMTLNAKRTSCSGQREAA--KTSALTILAS 778
Qy      2565 SFPAVILLSTATWLLALLSVNSDTLLPHYLTPATNCIOGPFIPLSYVLSKRVKALKLAC 2624
Db      779 SFLLILLVSASMLFGLLAVNHSIIAFHYLHAGCGQLAVLLPFCVLNADAAAMWPAIC 838
Qy      2625 -SRKEPPDALTKSTLTSSYNCPSPRADGRLYQ--PYGDSAGSLASTSGSKQP----- 2677
Db      839 LGRKAAPBEERAPRGLGPGAYNNTALFEBSGLRITLGASTVSSVSARSGRTODDSOR 898
Qy      2678 --SYI--PFLIRBESA-----LNPQGPFGIG-----DPSLFLBQDDQHDPTDS 2720
Db      899 GRBYLADNVLVIRGSAADHTDHLQAHAGPTLDVAMFRHDAGA-----DSBS 946
Qy      2721 DSDLSLEDQSGSYASTHSSESEEBEEREAAPFGEQGMDBLILPGARLPLHSTPYO 2780
Db      947 DSDLSLEEBESLISIPSESEEDNGTRGRFORPLCRAAQ-----SERLLTH--PQD 994
Qy      2781 GCGGPGKAPPG-----DRGT-----TAKSSGNGABERLREKDALSTRE 2821
Db      995 VQDNDLLSTVPALGECBAAPCALQITWSEERLGLDITSKAAANNQDPPAL--TSGD----E 1049
Qy      2822 GSLGPLPGSAPQHKGLKKKCLPTISEKSLRLP-----LEOCTSSRSSSA-- 2870
Db      1050 TSL-----GAGQRKGLIKR-----LQYELVPTQRCAPELSWCRAATLGHRAV 1094
Qy      2871 -----SESGRGPPEPPPPQSLQEOJN 2893
Db      1095 PAASYGRIVAGGTGSLGPAASYSSREOLD 1125

```

---

```

IUFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #ext_change 16-Feb-1997
C/Accession: A41087; B41087
R/Name(s): P.A.; Weber, U.; Onofrechuk, P.; Blessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A/Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A/Reference number: A41087; PMID:92069752; PMID:1959133
A/Accession: A41087
A/Molecule type: mRNA
A/Residues: 143-485;1279-5147 <MA>
A/Cross-references: GB:M80537
A/Accession: B41087
A/Molecule type: DNA
A/Residues: 1-142;487-1278 <MA2>
A/Cross-references: GB:M80537
A/Note: 1229-Gly and 1233-Ser were also found
C/genetics:
A/Gene: fat
A/Cross-references: FlyBase:FBgn001075
C/superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C/keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F/1-35/Domain: signal sequence #status predicted <SIG>
F/36-5147/Product: cadherin-related tumor suppressor #status predicted <EXT>
F/36-4583/Domain: extracellular #status predicted <EXT>
F/51-156/Domain: cadherin repeat homology <CR1>
F/159-270/Domain: cadherin repeat homology <CR2>
F/271-382/Domain: cadherin repeat homology <CR3>
F/390-494/Domain: cadherin repeat homology <CR4>
F/497-599/Domain: cadherin repeat homology <CR5>
F/602-708/Domain: cadherin repeat homology <CR6>
F/718-822/Domain: cadherin repeat homology <CR7>
F/831-942/Domain: cadherin repeat homology <CR8>
F/948-1049/Domain: cadherin repeat homology <CR9>
F/1052-1153/Domain: cadherin repeat homology <CR10>
F/1155-1278/Domain: cadherin repeat homology <CR11>
F/1281-1384/Domain: cadherin repeat homology <CR12>
F/1387-1489/Domain: cadherin repeat homology <CR13>
F/1492-1601/Domain: cadherin repeat homology <CR14>
F/1607-1713/Domain: cadherin repeat homology <CR15>
F/1717-1823/Domain: cadherin repeat homology <CR16>
F/1826-1922/Domain: cadherin repeat homology <CR17>
F/1925-2027/Domain: cadherin repeat homology <CR18>
F/2028-2167/Domain: cadherin repeat homology <CR19>
F/2169-2278/Domain: cadherin repeat homology <CR20>
F/2281-2384/Domain: cadherin repeat homology <CR21>
F/2387-2491/Domain: cadherin repeat homology <CR22>
F/2494-2596/Domain: cadherin repeat homology <CR23>
F/2599-2703/Domain: cadherin repeat homology <CR24>
F/2707-2810/Domain: cadherin repeat homology <CR25>
F/2813-2913/Domain: cadherin repeat homology <CR26>
F/2915-3013/Domain: cadherin repeat homology <CR27>
F/3014-3124/Domain: cadherin repeat homology <CR28>
F/3127-3229/Domain: cadherin repeat homology <CR29>
F/3232-3334/Domain: cadherin repeat homology <CR30>
F/3337-3439/Domain: cadherin repeat homology <CR31>
F/3442-3545/Domain: cadherin repeat homology <CR32>
F/3548-3651/Domain: cadherin repeat homology <CR33>
F/3654-3756/Domain: EGF homology <EG1>
F/3954-4010/Domain: EGF homology <EG2>
F/4017-4048/Domain: EGF homology <EG3>
F/4056-4089/Domain: EGF homology <EG4>
F/4096-4127/Domain: EGF homology <EG5>
F/4584-4609/Domain: transmembrane #status predicted <TM>
F/4610-5147/Domain: intracellular #status predicted <INT>

```

---

```

Query Match      10.0%; Score 1549.5; DB 1; Length 5147;
Best Local Similarity 22.8%; Pred. No. 3.3e-70;
Matches 553; Conservative 293; Mismatches 674; Indels 903; Gaps 71;

```

---

```

Qy      161 RSPESLGGRRKRVNTLPQFPSPYQATVPENQAGTVPASLRALDDPBGAGRLTYTM 220

```

Db 2258 QSTESQLVRLIDENDNSPVFDPKOYSASVAENASIGAMVLQVSAITDVEGANGRIRYSI 2317  
 QY 2221 DALFDRSNQFSLDPYGAVTTAB- 246  
 Db 2318 -VIGD--QNHDFSISEDTGVVRVAKUNLIRLSYLTRAEDCALBNPAGDTAELTINI 2374  
 QY 247 - 246  
 Db 2375 LDINDNRPTFLDSPYLARVMENTVPNGGYVLTVNAVDAOTPPPLNSQVRYFLKESGDLF 2434  
 QY 247 -LDRETSVFRVYTAODHGMPPRSALATLTILVTDNDHPVPEBOQ 292  
 Db 2435 RINASSGDIALKLPREBOSEKTYLLVAMDTGSPPLGTGIVRVEVODINDNDPVFELQ 2494  
 QY 2293 EYKESLRENLEVEYVLTVAATDGDAPPNANILYRLLEBSGSGSPSEVFEIDPRSGVIRTR 352  
 Db 2495 SYHATYRENLPSGTHVLTTPATDDBGLNAKLRNLL-GERMERFHIIDSTGEISIA 2550  
 QY 353 GPVDRREVESEYOLTVEASDQGRDPGPRSTTAAVFLSYEDNDNAPOFSEKRYVVOVREDV 412  
 Db 2551 TTLDRRETSVYHLTLMAQDSITE-PRASSVNLTISVDVNDNI PKFDSITTVNAVAPERI 2609  
 QY 413 TPBAPPLARVASRDGKSNVHYHYSIMSGNARQPTLDAOT- 453  
 Db 2610 SKGEFVGARALDLDDGENAVVHYTI--SGRDQHYFDINTKTGVVSTKLELTKTKSTD 2666  
 QY 454 -GALD- 457  
 Db 2667 DLTTTVISAMDQEGSLSKAELTYLRPELPPTPAVANSHTMSEVBRGKMTTKV 2726  
 QY 458 -VVSP- 480  
 Db 2727 SATSPKGLVGIKIRYAIAGIMGDSLREVDPMGGLSVQDGLDELYHLYEIMWBAADG 2786  
 QY 481 RPLSNVSGLVYQVLDINDNAPIPVSTPQATVLESPGLYVAVQALDADAQNAUL 540  
 Db 2787 TPLSRST-LITNAVTDANANAPMEQLITYAELBESPPOLIAYKASDRSGNGNANV 2845  
 QY 541 EYRLAGVGHDPFTIN-NGTGMISVAALDREVEDFPSFVEARDHGTPLATASASVY 599  
 Db 2846 IYHLQ--NDFDGFELTESGEITYRRLDREHIGDAFVEAVDQVPHLTGTASVYLH 2902  
 QY 600 VLDVNDNPPFTQPEYVRLMEDNAGTSVYVTSANDRD--AHSVITYQITSGANTNRS 657  
 Db 2903 LLDKNNPPKPTR-LFSLANTENAKISFVIRVTSDDLGDANANASYSF-SEMPGKFR 2960  
 QY 658 ITSOGSGVSLALPLDYKLERQVYLVATSDGRODTAQIVNAVVDANHRVPOSHY 717  
 Db 2961 IEPQSGN-ITVAGHLDRQODEYILKVASDAMRAETPTITTIQDQNNAPFEHSFY 3018  
 QY 718 TVANVEDRPAGTVVLISATDSD-TGENARITYFMEBDSIPQFRIDADTG- 765  
 Db 3019 SPSPEPQOSIALVGOIATDRDKQGNVSISSLOQSPMFSIDPATGSEPSKAVRFK 3078  
 QY 766 -AVTQA- 771  
 Db 3079 HSOYVSPENMYALTVALTNGKPEPLXSECLVININIVDAHNPKEFOAEYLAPLPODAV 3138  
 QY 772 -ELDY- 784  
 Db 3139 RQGRIVRHANDKODLTENEDYSMTFNLSIFSVGRHDGMITLVKPIOVPPNTRJELV 3198  
 QY 785 ITARDNGIPKSDPTTYLEILVNDVNDNAPQFLRDSYSGSYVEDPPTSVLQISATDRS 844  
 Db 3199 VRATDRGVPQSDETRVIIVTGENMDTPRFSVNSQVIVPENPEVSTILVTGATDDDT 3258  
 QY 845 GLANRVTYTOQGDGDGDPVVESTSGIVRKLRLDBENVAQVYLRVAVADKMPKATP 904  
 Db 3259 GPNGMKLYSISGGERO-DESVBERTGVIYQOQLDLDLIOEYHMLNTVVDLGHPLSSV 3317  
 QY 905 MEVTVTVLDVNDNPPVEODEFVFEVENSPIGLAVARVATTDDEGTNA- 954  
 Db 3318 AMULTIILTDVNDNPPVRNKHGYHCTIPENKFGVGFVQAHATDKDSKNAIIRHAPLPSG 3377

QY 955 - 954  
 Db 3378 PDRHEPIMQSNGTISSAVSPDYERRIYTLQIKANPDSSMESTANLYVHVLGVNEFP 3437  
 QY 955 - 954  
 Db 3438 QFLQVPHFDETSAGVTRVGAVOATDKDGEDGRVYVLLVGSSNDKFRIDTWTGLY 3497  
 QY 955 - 954  
 Db 3498 VARHLDRETQNEVLTMAKNYGSIRGNDTDEAQYIISIQDNDEPEIKHYTSTISBA 3557  
 QY 955 -QIMQYVEGNIPEVFOIDIPSGELTALVDLYEDREPY 992  
 Db 3558 AVGKNTTVKALIDVARTQNNQPSYSIINGLKQSPKIDVGTGHSITSRIDRETSY 3617  
 QY 993 VLVIOATSA--PLVSRAVHRLLRDNDPVL--GNFEILLNNYVTNRSSTPGCAI 1046  
 Db 3618 NIVIGALIDTGLPQTSATVHIELEDVNDGPTPTPEG--LNGYISENEPA--GTSI 3670  
 QY 1047 GRVPAHDPI--SDLSYSPERGNELSVILNASTG- 1080  
 Db 3671 MTLMASDPDLPRNGGPFYTLQIGKHKSWLVDRNSGVVRSSTSPDREMTPLYALIEVE 3730  
 QY 1081 - 1080  
 Db 3731 DSGKPKQSQHLITTVLDQNDNPSTTSLHIAVSLFNGDLPSNVKLVADVRNDIDYGD 3790  
 QY 1081 -EKLRSALD--NNRPLEAIMSVLSDGYH-SYTAOCALRVITI 1120  
 Db 3791 YRCRLQKNPAQOLALIPRACDLITTSHTTPIASVFTGDKGKGDVSSVSAFOSF 3850  
 QY 1121 TDEMTHSTTLRLBMSPERFLPSLGLPIQVAATTLTPPHVYVFNQRTDAPGHI 1180  
 Db 3851 NNKETLANSVIVRNMTANHFLANHTRPILBETIKRMSN-DEEVILYSL--LEGGGNS 3906  
 QY 1181 LNVSL-----SVGQPPGPGPPPLPSBLOERLYNLSLTJASQVRLPEDDNI 1231  
 Db 3907 TNLQILMAVRLAKTSYQD-----KYLIERLEKKSARSELQKEYIVGYEBC 3954  
 QY 1232 CLREPCENRCVSLRFPSSAPFIASSS--VLRPHIVYGLKCRCPPTGD- 1283  
 Db 3955 SBDVCENGVCSATMRLIDAHSPVODSPALVLSGP-RVHDVYSQCTSGSPGSCSR 4013  
 QY 1284 -YCETB-VDLCYSRPCGPHGRC-BSRSG 1308  
 Db 4014 QDPCLPNCHSOYQCRRLASDPOCMCPANRDGKHEKERSDYCYKPCANGSCQRSPPG 4073  
 QY 1309 -GYTCLRDGYTGEHCEVSABSGRCTPGVCXNGGTCVNLVGGFKDCPSGDFEKEYQV 1367  
 Db 4074 SSYFCLCRGFRGNQCE--SVBDSRPNPCLHGLCVS-LKGYKNCCTPRGYGR-HCER 4129  
 QY 1368 TTRSPAHSPITPRGLRQRFHTLALSPATKRDGLLYNGRPN--KIDPALAVIYQO 1425  
 Db 4130 FSYGQPLSYMTFPLADVTTN-DISIVFATYTPNSLLNYNMOGSGRDFLALIBVHR 4188  
 QY 1426 VOLTFEAGSTTVSPFPVG-GVSDQOMHTVOLKTYNKKLLAQTGLPQSPBOKVAVTV 1484  
 Db 4189 AY--PSSGAKRTAISVIVAGRLADOGMHKV--TATRNG--RWMSLSV 4230  
 QY 1485 DGDGTVALRFGSVIGANVCAAQ-----GTQGSKKSJDLTGPLLLGV- 1535  
 Db 4231 AKCADSGDVCTECLPDSSCYADVEVPVGTILFNQO-----PLMIGJLSSADPILEPBG 4284  
 QY 1536 PYRMQFVGCMMNLQVDSRHIMADPIANGVTPGPACQVCD-----SNTCHN-GGTV 1590  
 Db 4285 QVHSDDLVGCLHSVHIGGALNLSPLQOKGJLAEG--NRQACQPALAERCGFRAGQCI 4342  
 QY 1591 NQMDAFSCEC-----PLGFGKSCAQ--EMANPQHFLG--SSIVAM-- 1627  
 Db 4343 DRKSSSLCCGGHLAGSDCSDSLPITLGBGAFVEFRIBSEIYRBMQLUNLNNSKAMLD 4402



QY 1628 -----HGLSLPISQFWLTLIMRTROAGVLLQATIRGRSTTLQREGRHW 1674  
 DB 4403 NOOMERRAVNSFSTASQIYEAKRMISMLFRFKDKQOQILYATNOMFT-SLSLRGRLV 4461  
 QY 1675 -----LSVEGTGLQASLSLLEFGRANDGDMHHAQALGASGGPGRHALLSFVGGORA-- 1726  
 DB 4462 YSKQHLTINMTVOERSTL-----NDGKMHVSL-----FSEKSLR 4498  
 QY 1727 ---EG-NLGRHLH-GLH-----LSNITVGIIPGAGVARGKRCLOGVRVSDTPEGV 1774  
 DB 4499 LAYDGOVQDELDIAGVHDFLDEYLTILNVG-----BAFVGLANVTVNNELQPL 4549  
 QY 1775 N---SLDPS---HGESINVEQC 1791  
 DB 4550 NGSSTIFPEYRYHCK--IRSGC 4569

## RESULT 5

T30213  
 G-cadherin - sea urchin (lytechinus variegatus)  
 C:Species: Lytechinus variegatus (variegated urchin)  
 C:date: 22-Oct-1999 #sequence\_rev1999 #text\_change 22-Oct-1999  
 C:Accession: T30213  
 R:Miller, J.R.; McClay, D.R.  
 Dev. Biol. 192, 323-339, 1997  
 A:Title: Characterization of the role of cadherin in regulating cell adhesion during sea  
 A:Reference number: Z20780; MID:98104238; PMID:9441671  
 A:Accession: T30213  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2809 <M11>  
 A:Cross-references: EMBL:U34823; MID:g2982186; PID:g2982187; PID:AAOC6341.1

Query Match 9.7%; Score 1508.5; DB 2; Length 2809;  
 Best Local Similarity 29.0%; Pred. No. 1.7e-68;  
 Matches 538; Conservative 252; Mismatches 757; Indels 311; Gaps 87;  
 QY 147 SCKLAQBPGLRAGE-----RSPESLGRKRNNTAPQOPPSYQATVPENOPAGTP 199  
 DB 938 SCIVASDSIRSRVDVDFVRSPPDT-----TYQVPTSPNTPYQVTINEMTTIG-D 988  
 QY 200 VASLRAIDPDEBAGLEFYTMDALPDSR-----SNQFSL-----DPYGAVTAAELDRPT 251  
 DB 989 VILLVSV-----GAQEVGPDIVKSRSTNSDGTFTLTDDNLQSDLTVEKLDPEA 1042  
 QY 252 KSTHFRVTAQDHGMRRSALAT-LTILVTDNHDHVPBEOQYKSLKLENLEVGVE--V 308  
 DB 1043 VQEVLELTVTAQYVGVQBFIAQPTVLTIEDFDVAPKRPATVPGNIENQPVNSDPI 1102  
 QY 309 LTVKATDGAAPPNANILYLLRSGSGSPSEVBEIDPRSGVIRTRGVDRREVE--SYOLT 366  
 DB 1103 IYLQATIDGVTPE---YSQISYSPNABSDSIDETGEIFALKVFDREALEDSIALT 1158  
 QY 367 VEASPOGRDPRGRSTTAATFASVEDNDNAPQFSEKRYVVOYREDEVTPAATLRTASRV 426  
 DB 1159 VASD-----GNODLATYFTITVDENDEPERN-GTSPFVLEDEGKMGYDGTATADD 1212  
 QY 427 DKGSNAVNYHYSIMGNARQFIYDAQTGALDVVSPLDYETKEKYLRYAODGGRPLSN 486  
 DB 1213 D--ISEVLEYPIISGNREGAFVYDAEGTIRKAGVLDYEAFTSYELQSVVNDG-----KN 1265  
 QY 487 V-SGLVTVQVLDINDNAPIFVSTPQATYLE--SVPLGLVLAHQAI--DADAGNARLE 541  
 DB 1266 VATTVTVTINLVNNDVAPQPDQASVASYIEBDNSLPRILLSVATDQDADAVDA--VV 1324  
 QY 542 YRLAGVGHDPFTTINNGTGMISVAELDEEVDYFSFGVEARDHGPALRTASASVYVL 601  
 DB 1325 YELVGTGACTITLISQITNTLTQALDREELPTTNLAVAATDNGNLSTVYDVTTIYE 1384  
 QY 602 DVNDNNPFTQPEYTVRLNEDAAVGTSVTVASVDRDAHVTITTYQITSGNTNRSPTSQ 661  
 DB 1385 DINDNAPVPPDQEVGSVEENRPPPTPVVAVAABDVTADDLMTSPPTSP--DENINSQ 1442

QY 662 SGGGLVSLALPLDYLT-ERQYVLAATASDGTRODTAQIVANVTANTHPRVFOSSHV-TV 719  
 DB 1443 T--GOITAKQDREPTPESEYEIEVOATDGVNTASTATVTSIDVDNDPSPSEDEVYDA 1500  
 QY 720 NNNBRPAGTTVVLISANDEVDGERARITYPMBSDIPQ--PIIDDT-----GAVTQAL 773  
 DB 1501 SVLETEPIGTITTVQALIDPDVDFRDAVEFSINSQDPDELFFIYVDAATLQGLQVNNKL 1560  
 QY 774 DYEDQVS--YTLATYARDNGIIPQKS---DTYALILVNDVNDNAPQFLADSQSGSYEDV 828  
 DB 1561 DETLATNREFLTVAVTDSQGTDSGRPETAVKLIYENANDLAAVFQDQYDGAVERSEA 1620  
 QY 829 PPTTSV--LQISATRD--SGLNGRVFTYQGGDGDGDFIVESTSGIVRTLRRLREN 883  
 DB 1621 TVNSQVGTFSATGDEPRGATLN---YIIDPTDPGQFAINNQQLT-VASPLDRPT 1675  
 QY 884 VAQVYLRAVADKGMPPARTMEVTVTVLDVNDNPPVEQDEPDVPEENSTI--GLAVA 941  
 DB 1676 VASVELKAYAVDNGPSPSGTATVAVTVTVNDVTPHFAQD--YNPSVEE-GEIENAVEV 1733  
 QY 942 RVTATDPDE-GTNAQIMYQIVEGNIPREVOLDIF-----SGEL---TALVDLYE 987  
 DB 1734 SVEAVDDDDPPSGPPLNVA-----PQNDWTFPDIGLGTSGISIRVSTYGLIURE 1789  
 QY 988 DRP--EYLVITQATAP--LVSRATVHRLDRNDNPVLANFELFNNTYNNRSSSPG 1043  
 DB 1790 THPYEDIVFLAIVGTPEALVTGTQTLTMDVNDNPVATIKOILVYSRGN---IPT 1845  
 QY 1044 GAIQVPAHDDDISLTYSPFERGNEI-SYVLNASTGEBLSRLDNNRPLEAI-MEVL 1101  
 DB 1846 TEVGAVGVEDPDILDKTY--EAVGELPDFOLSDTDITMA---EGTPAGYEMDIR 1899  
 QY 1102 VSDG--VHSVTAQCALRTYITITDEMLTSHITRLSDMSPERFLSP-----LGLIF 1149  
 DB 1900 VSDQTSVSVSVTVIVYKQIPRAVRESSGSIIRSGTAAELITPDSGVSNDRLKYL 1959  
 QY 1150 IQAVATLATPPDHVNVN--ORDTDAP--GSHLVNLSVGGPPGGGPPPLPSDL 1205  
 DB 1960 ANAVGAQLAN-FDIFSVLVNVGMERTVDIRYAAH-----GSPYPAQOL 2002  
 QY 1206 QERLYNRSLTLTAISAQ-----VLPRDNICLRE-PCRNVMGCVLARRDSSAP 1254  
 DB 2003 D-----LAATSVSREIEGLSLAQIPVD--LCVENVCBS--SCTNVLVVDPIPT 2049  
 QY 1255 FIASSVLEFPIHPVGLRCRPPGFTGYCETVEDLYSRPCGPHGRCSRREGYTCLC 1314  
 DB 2050 VVDSGTALVAITSV--LEAQ-----ICGANTVAP----- 2078  
 QY 1315 RDGTTGHECHVASAGRCTPGVCXNGGTCTNLLVGGFKDCPSGDPEKPYCQVTRSPA 1374  
 DB 2079 -----GNQSDSPCLNNGTCEDVHGCTVACTPY-LFDGPNCOQTRKSP-E 2121  
 QY 1375 HSFITPGLRQRFHPTLALSPATKERDGLLYNRF-----NEKNDPFALEVIGQVOLT 1429  
 DB 2122 NGVASFSTLRCEBTSISIEFTIVSSGTLXNBPITPTGDDPDIIMLLLEIGKALYT 2181  
 QY 1430 FSAEISTT-----VSPFVPGVSDQGMHTVOLKYNNKPLAQYGLPOGSBOKVAVVT 1483  
 DB 2182 INLSTSTSDNVLBAAPTDQTQNDNMRHIDV-YNN-----GRFEMT 2225  
 QY 1484 VDCGDTGV--ALRFGVLYNYSQAQGTQGSK--KSLDITGLLIGGV-PDLPESEPV 1537  
 DB 2226 VDRG-MGFPAETSSSSTLDDSSCRVNGTTPGKNNLRFNLNHTPLFLGMSADVDVVP- 2283  
 QY 1538 RMQPFVGMRLQVDSRHIDAPFIANGVTGPGPAKKNVC--DSN-----TCNNGGTGNQW 1593  
 DB 2284 --SGFDGCIKULVSDGLYDGTGTSKSAEGCPRIQDCTDNGMPVCNN-GTCEADI 2340  
 QY 1594 DAFSCECPLEGKGSQAQEMANPQHLGSSLVANHGSLPI---SOPWYLSIMRTQAD 1650  
 DB 2341 DSFICICPPGNGLTQVLEL-TPYDAIISYITIELDSSLYDARSNNYQIMVTRQEN 2399  
 QY 1651 GVLQALTRGNS--TITQLRBEHVMLSVBGTGLQASLSLLEPRANDGMH-----H 1701

```

Db      2400 G-LIWSISSANTYEYIMEMVQGLKADWH-LGDKPVSVMNFSINDGMAHINFDKXD 2457
Qy      1702 AQAALASGGPGRHAILSPDVGQGRABGNLGRHLGHL-SNITVGGIPGAGAVGAPFG 1760
Db      2458 SVVTIKIDGGGGVKEIO-----NRESQYSGLDVENSIVIGAFVDVNTVTDPMG 2507
Qy      1761 CLQGVASDPT---PEGVNSLDPSHGESINVEQCSLPDPCSNPCPM-SYCSNMDSYS 1816
Db      2508 CMMDPRIINHHFLMGEGNDNAVA-TKSAGTBGP-SDVCDSDPCGSLIVCTDYRFPTE 2565
Qy      1817 GSCDPEGY-----GDCTNVCIL-NPCHQSVCTRRKSPAPHGTCCEPNYIAPYCE 1867
Db      2566 CLCPBGEVEVEDPDCTMAIIDCVPMPCANGTCV--GDPTGTCDCPSGGYGRCE 2621

RESULT 6
T00021
DN-cadherin - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 17-Nov-2000
C/Accession: T00021
R/twat, Y.; Ueni, T.; Hirano, S.; Steward, R.; Takeichi, M.; Uemura, T.
Neuron 19, 77-89, 1997
A/Title: Axon patterning requires DN-cadherin, a novel neuronal adhesion receptor, in th
A/Reference number: Z14058; MUID:97388431; PMID:9247265
A/Accession: T00021
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-3097 <IWA>
A/Cross-references: EMBL:AB002397; NID:g2381491; PIDN:BAA22151.1; PID:g2381492
A/Experimental source: whole embryo and adult head
C/Genetics:
A/Cross-references: FlyBase:FBgn0015609
A/Map position: 36D
A/Note: I(2)3cDA
C/Superfamily: unassigned EGF-related proteins; EGF homology
F/2346-2377/Domain: EGF homology <EGF>
F/2869-2902/Domain: EGF homology <EGF>

Query Match      9.1%; Score 1421.5; DB 2; Length 3097;
Best Local Similarity 24.7%; Pred. No. 5.5e-64; Indels 533; Gaps 72;
Matches 514; Conservative 286; Mismatches 747;

Qy      179  POFQPPS-YOATVPENQAGTPVVASLRALIDPDEGAGRLTYMDALFDSRNOFFSLDPV 237
Db      966  PVFQDSTYTPKVEBAPNSPVIKVATDEKGVNQVYSIQGNQGTGK-FTVDEE 1024
Qy      238  TGAVTTAEBLDRRTKSTHVRVT--AODHGMFRSALATTITLVTDNDHDPVPEQOEYK 295
Db      1025  TGRVSTKRVFDRREGDDKPFVSVTKATDQGPDLSEGVCSFTVBLITVDNDPPLFDRQKYV 1084
Qy      296  ESLRENLEVEYVLTATATGDAPPNANILYRLLEGSGSPSEVFELIDPRSGYIRTPGPV 355
Db      1085  ENVAQDSIGTINILRVASBEDADNNGAIVYSLTAPNPDLBEFELQASGSIYVKKPL 1144
Qy      356  DREVESYQUTVEASDQGRDPGRSTAAVFLSYEDNDNAPQSEKRY-VVOVREDVTP 414
Db      1145  DR---EYTKLEANAQDGYR--PLSRVTEQIDVVDANPNPVMWDHIVYPIVYKEMPV 1199
Qy      415  GAPTATATASDRDKGSNAVHYISMSG-----NARGQFYL-----DAQTGA-LDVYSPLD 463
Db      1200  GGRKVASIKASSGEG-NPTVYRLMPGSTAQTNNKFFYLLQQRPNEDTVAADIKVNHPLD 1258
Qy      464  YETTKETLRLVRAQDGRPLSNVSGLVTVQVLDINDNAPLIVSTPQATVLESVPAGYL 523
Db      1259  YESIKETNLTRVANNQAQLAS-BATYITMLEVDNDEIPLFTRE-QEYVLSEEPITGK 1316
Qy      524  VLVQAQIDADAG-----DNARLEKRLACVGHDPPTINGTGMISVAELDREE 572
Db      1317  VTQNALIDKQGFENNQYVYIYDSARNE-----GKEF-FEINLQSGEIFTKVVFDRK 1369
Qy      573  VDFPSFEVEARDHGTPLATASAS-----VSATVLDVVDNNTFTPOPEYTRINDDAA 624

```

```

Db      1370  KCAVALLEVEARD-GAPSAEPNNGSNVTKFRIIGLAKNDNPFPYFDSLYAEADENED 1428
Qy      625  VGTSVVTAANDRQSHVITTOITSGNTRNRSISQSGGGLVSLALPDYLERQYVLA 684
Db      1429  IOHTVITVAKDHSSIRIRIISGNTGAPAVNMGA--TVAGALDIESTRRIELR 1486
Qy      685  VYASGTDQDTQAIYVNVYTDANTHRPVQSHYTVNVE-----D 724
Db      1487  LAASNLKENTYTVIIHKQVNDNPVFERPFRYQITEBDRNLKRYLQVATDGDKD 1546
Qy      725  RP----- 726
Db      1547  RQNTIVYELTQGIDPDNPANSKPDINRTGELFVLKELDRDQPNRQPMREVEADQEG 1606
Qy      727  -----AGTVVLISATDED-----TGENA 745
Db      1607  GEGLVYADVQVNLKINDNAPIFQGYVFGAVTENGKAGVMTMTAVDIDDPNAGSNA 1666
Qy      746  RITYEMEDI-----PQFRIDADTGAVTQA----- 771
Db      1667  RLVTSLKKNVIEETGSPILFEIBPTGVIKTAVCCIDBERFPDYSIQVAMDGGLKGTG 1726
Qy      772  ----- 771
Db      1727  TASIRVKQINDMPQPTKDEWETVEDTGTALPEMPLITVTHDEDETNNKQYKVIDNS 1786
Qy      772  -----BLDYEDQV-----SYTLATFARDNGIPQKSDT-----TY 800
Db      1787  GYGAOKTWRNRNDGTGSLIKIVQPLDYEDQLSNFRRIQVNDNGENDNKKHYAVYSW 1846
Qy      801  LEILVNDVNDNAPQFLRDSYQGYVEDVPPFTSVLQISATDRDSGLNGRVFTYFGGDDG 860
Db      1847  VVVKLRDINDNKHFERANVAVSEVFEETKVEKFKATLDPDQGGKSVSYISRSBR 1906
Qy      861  DDDPIVSESTSGVRLRLDRPENVAQVYLRANVUKMPAPATPMENVYTVTVDVNDNPV 920
Db      1907  QROFAL-NONGSVTLTQSLDRVVRHGVKILATIDGSPKTAATLTLYIVQDINDNAK 1965
Qy      921  PQDEEDVVEENSPIGLAVATATDPDEGTAQIMYQIVGINPEVQDLIFSGE--- 977
Db      1966  FLKDRPVLPEHVPP--HKVSEILATDDDRKS-----NGPFGRLDPSADDIR 2015
Qy      978  -----LVALVDLYEDRPYV-LYIQATSNP-LVSRATVAVRLID 1015
Db      2016  ASFKYEQDQKANGDGMVAVISSLSRFDREQKRYMIPYIKDGPAMGTSTLTLYIGD 2075
Qy      1016  RNDNPVYLGNEFELFNNTVYTRNSSSPGGALGRVPAHP--DISDSLY-----SPE 1065
Db      2076  VNDNKKQPSKQIFVYNY---QGQSPDTPICKRVYVYLDMDLDPKKFTWEMAEHPRK 2131
Qy      1066  RGNELSLVILNASTGE-----LKLRLADNRRPLEAISMVLYSDGVH-SVTAQCALRYT 1118
Db      2132  LDEDSGMTVRAQTRGRLHAFKYDRKHGTQDIPAVTVVARELIPHEAVNNGSVRLS 2191
Qy      1119  ITDEMLTHSTLRLEDSPPERFLSPLGLFQAVATATATPPDVVYVNVQRODABEG 1178
Db      2192  GISDEDFIRVNNYKQOSMSRSK-----MDRFQDKLADLINTREAVYDFSVLKRKP-- 2244
Qy      1179  HILNYSLSV-GQP---PGGQGPPLPSEDLQERLYLRSLTALASQGLVLPFDNICKR 1234
Db      2245  PLTDVRFASGSPYKPYRANIGVLMHREIKQVGIN--ITMGIDBCL-YENQMC-E 2299
Qy      1235  EPCENYRCVSVLRDSSAPFIASSSVLFRPIHPVGLGRCPGFTGDCETEVDLCYS 1294
Db      2300  GSCNLSLEI-----SLPYYVNNAN-----KTALVGRVDTIAD----- 2332
Qy      1295  RQGHGRGRSRSGGTCLCRGTYGEGHCEVARSGRCTPGVCKNGCTVNLVVGFKCD 1354
Db      2333  -----C-----TCGARFTKPESC-----RTTP--CHNGKRCVDRFGP-HCS 2367
Qy      1355  CSPGDEKPYCQVTRSPAFSFTFRGLRQRFHTLALSPATKERDGLLYNGHFNEKH 1414

```

Db 2368 CPVQ-YTGERCCQTRSFSGNGMAYPLEMCDESHLSLEFTRRDPGLIYNGPIVPE 2426  
 Qy 1415 -----DFVALEVIOEQVQLTFSGESTYTVSPVPGVSDGOMHTVOLKYNRPDLGQ 1467  
 Db 2427 RDELISDPIALERGVRLIIDGSSGLBELRVKTKLIDGEMRIDLFPDTSIR-- 2484  
 Qy 1468 TGLPGPSRQKAVVTVVDC-----DTGVALRPGSVLNGSCAAQGTGSGKSLDL 1519  
 Db 2485 -----NVVDFCKSAELAEEMEDGTPPEFDM-----SCQARQIIPFNEVLNV 2526  
 Qy 1520 TGPLLGGVPLPESPRYM-----RQVGCGRNLOVDSRIIDMADFIANNGTYPG 1570  
 Db 2527 NAPIVGGI--YRQPDOSLYRMTHTPTAKGDFGCTRLVHNSKLYDLAHPGLSNVSAG 2584  
 Qy 1571 CPKRNVC-----DSNTCHNGTCVANOQMDAFSCCEPLFGSGKSCAQMANKPFLSSLYA 1626  
 Db 2585 CPQTEVCAQETETARCMWHGNCVCSLSBARCHCRPGWTGPAC-NIPITPTFKAQSYK 2643  
 Qy 1627 WHGLSLPISQFWYLSLMTTRQADGVLLQATR-GRSTITLQLRGAV--MLSVETGLQ 1683  
 Db 2644 YALSEPRPFTQVOLFRTREEREGELFRVSDQHNREYGLEIKDGHLPFRNLNLSLRT 2703  
 Qy 1684 ASSTLEPGRANDGMHHAQLLGLASGGRHALISPDVQGRABGN-----LGR 1733  
 Db 2704 EKDLMLAIIVNDGQMHVVKVRYS-----ATLELDGGRNRYETEPVGHQMLLVDK 2759  
 Qy 1734 LHGLLSNITVGGIPAGVA-----RGF-----RGCLQGVAVS-----DTPGVNSL 1777  
 Db 2760 QEGVY-----AGKALEYGVRTFRVYADYQKTCDDILYEGKHLPLPAMNG- 2806  
 Qy 1778 DSHGSEST--NVEGCSLPDPCDSNCPRANSYCSNDSDYSQSC-----DPGYGNC 1828  
 Db 2807 -TOMQATMANRLERKGCSCNKCNSVTCPDPECVDLNMVVECTGCBGRIMSPDKGMD 2865  
 Qy 1829 TNYCDLNPCHQSVCTR-KPSAPHYTCECPENYIGPICE 1867  
 Db 2866 RNECLDMFCMNGATCINLEPRL-RYRCICPDGFGENCE 2903  
 RESULT 7  
 T00252  
 MEGP1 protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00252  
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.  
 Genomic 51, 27-34, 1998  
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs  
 A:Reference number: Z14126; MUID:98360089; PMID:963030  
 A:Accession: T00252  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-4351 <NAK>  
 A:Cross-references: EMBL:AB011527; NID:G3449285; PIND:BA032458.1; PID:G3449286  
 A:Experimental source: Brain  
 C:Genetic8:  
 A:Gene: MEGP1  
 C:Superfamily: rat MEGP1 protein; cadherin repeat homology; EGF homology; laminin G repe  
 F:3798-3949/Domain: laminin G repeat homology <LGR>  
 F:3953-3985/Domain: EGF homology <EGF>  
 F:3992-4023/Domain: EGF homology <EGF1>  
 Query Match 8.2%; Score 1275; DB 2; Length 4351;  
 Best Local Similarity 24.0%; Pred. No. 2,8e-56;  
 Matches 534; Conservative 271; Mismatches 830; Indels 588; Gaps 77;  
 Qy 179 POFQPSYQATVPENQAPAGTPVASIRAIIDPDEGRAGLETMDALFDSRSNOFSLDPT 238  
 Db 2377 PKFRPQYKAVNSRLATGSHLVKVALDLPDIDGTSRLY--LILSGQDRHFSINSTS 2433  
 Qy 229 GAVTTAEELDEKTSYFRVTAODHGMRRSALATLLIIVTPTDHDVPEQGEYKSL 298  
 Db 2434 GIIISFNLCKKOLDSSTNLKVGASD-GVFR--ATPVYIINTNANKSPFQONVYAEEL 2490

Qy 299 RENLEVEVLTVRATDGDAPNNILYRLLEGSGSGSPSEVEFIDPRSGVIRTRGPVDR 358  
 Db 2491 AENAKVGKVELIALIDDSGPFYGVDTIINKLAG---EFFINFR-GQITTLQKDR 2546  
 Qy 359 EVESYQLTVEASDQGRDPFRSTAAVFLSVEDNDNAPQSEKRYVQVVEDYTPGAV 418  
 Db 2547 --NSTERVIALIKWARRDGGKVAFCVXKILTDENDNAPQKASGYTVISNVRSDPI 2604  
 Qy 419 LRVASDRCKSNMVAHYSIMS----- 440  
 Db 2605 IQVLAYDDEGRNADVTVSDSTEDLAEILIEVNPPTGVVKSGSLGLENRAVDENIKA 2664  
 Qy 441 ----- 440  
 Db 2665 QDGGPRHMDSLVPRLOVPRBEIPLPKRSEPLTFPSAPEDPREGSEIGSVKAAVADPTI 2724  
 Qy 441 -----GNARQPYLDAQTGALDVSPLDYETTKETLRYRA-----ODGARPLSN 486  
 Db 2725 YSLVQGTTPESNSDVSFSLDDQTVLKVRKAMDEHSTKMYQIDIMAHCPHEDTLVSLVS 2784  
 Qy 487 VSGLVTVQVLDINNAPLIFVSTPQATVLESVPLGYLVLYQALDADGNARLEYRLA- 545  
 Db 2785 VS-----IQVEDVNDNRVPFEADPYKAPLTENMPGGITVLIQVTANDQDTGSDGVYSRLSV 2840  
 Qy 546 --GVGHDPPTINNNGTGMISVAELDEBEVDYFSGYEARDHG--TPALTAGASVTVLD 602  
 Db 2841 EPGSNHHELFVADSESGHITTLQELDCETQVTFYVVAIPHGQITQLSSQALVESTLD 2900  
 Qy 603 VNDNPPFTQPEYTVRLNBDAAVGTSVYVSAVDRDA--HSVITYOITSQNTNRPESIT 659  
 Db 2901 ENDNPPRPSADYGSVVENNEPGLVATLKTLDADVSDQRQVTCVITTEDPILGQFIS 2960  
 Qy 660 SQSGGLVSLALPLDYLEGRQYLAATASDGTBDTQIIVNVTADNTHRPVPSHNTV 719  
 Db 2961 QVDEMRISRKTLDRHIAIYLLRVASDQKQASVVEFVVDINDNSQCSQLLYTG 3020  
 Qy 720 NVNEDRPAGTIVVLSATDEDTGENARITVFM--DSIPQRIADATGAVTQAELEDDQ 778  
 Db 3021 KYREDVTPGHILKVASAIDVDMQNAQITYSLSHPGAQBFRLDHTGELTILVYLDNRK 3080  
 Qy 779 VSYTLAITARDNGIPQKSDTTYLEILVNDVNDNAQPLRDSYQSGSYVEDVPTSVLQIS 838  
 Db 3081 DVNMLVAKATDGG--GQSCQAEVTLHIEDVNDNAPREFPSCHDVAIVEDNTVTKPVAVF 3138  
 Qy 839 ATDRDSGLNGVPTFGQGDGDGFVIESGIVRTLRRLRENVAGYVLAAYDKM 898  
 Db 3139 ARDPDQANMQVYSLT--DSADQSFSDATSGVIRLEKPLQVRASSAVELTVRASDLGT 3196  
 Qy 899 P-PARTMEVTVTVLVNDNPPVEQDEPFVFEENSPIGLAVARV--TATDP-DEGTNAQ 955  
 Db 3197 PIPSTIGTVTVSVAGLEDVLPFLANHEHQVEDEAPIDMEVHLATLIRPGSEKQ-- 3254  
 Qy 956 IMYQIVENQIPEVFOQDIFSGELTALVDLYEDRPEVLYVIA--TSAPLVSRATYHVR 1012  
 Db 3255 --YHITGNEQCKRFLAHGTGLYVNGSLDPEETPKYFLGISCRKSSSSLDVTTIYIN 3312  
 Qy 1013 LLDENDNPP-----VLGNFELLFNMYNNRSSPFGAIGVPRANDP--ISDSLVSFER 1066  
 Db 3313 VTDVNEHHPRTHDLYTVRVLNNAV-----GDVILTVSASDDGPNVSAITYSVLVG 3364  
 Qy 1067 GNBLSLVILNASTGELKLSRALDNNRPLEAIMSVLSDG-----VHSYTAQCALAVTITD 1122  
 Db 3365 GNGGHTINPKKGLQVAKALDMEQRPYSILRARADSQGPLHEBT--EVAVEVDVND 3423  
 Qy 1123 EMLTHSTTLRLSDMSPPRFLSPLGLFIQAVATLATPPDHVVVFNQORDTAPFGHIIN 1182  
 Db 3424 -----NPRPFQ-----LWYSTSVQENSPIG--IKV 3447  
 Qy 1183 VSLSVGPQPGGGGPPPL-----PSBDLQRVLYNBSLTAISQORVLPFDNITCLNEPC 1237  
 Db 3448 LQTLIDDPDFQNGPFPRFRITGNTGSVFRVVPDGMVLTNALS-----KKAR 3496

```

QY 1238 ENYKCVSLRFPSSAPFIASSSVLPRIHPVGLRCRCP-----G 1279
DB 3497 EWTQMLIEV---SDSGLPPLSSSTLVRQVRE---QGRYPPSTPLPLEISTTKBBERGG 3550
QY 1280 FTGDYCTEVD---LCYSRPGPHGRCSRBSGGTTLCDYGTGHECEVSANSRGCTPG 1335
DB 3551 MIGIHAITDDBDPPDPTLTYS---LEOBGGIDRFTYTGASGK----- 3588
QY 1336 VCKNKGTCVNLVGVFGCDGSPGDEKPYCQVTRSPFASFTPRGLRQRPHPLTALS 1394
DB 3589 -----ILASQGLPHGRISFNTVVS 3607
QY 1395 FATEKRDGLLLNGRFNEKIDPVALLEYIQRQVOLTFSAGSSTTVSPFVGVSDDQWMT 1454
DB 3608 -----DGTFTTTGVVHVHMHMBEVPRQAVMLGF-----HQLTEBELVSDHMRN 3652
QY 1455 VOLKTYKPLLGOTGLPQSPSEQKAVVVTYDCCDTGVALPFGSVLGNYSQAQGTQGGSK 1514
DB 3653 LQRFSLMLDVKRAMIHLSIQPAEVTAGVD---VLTVF-----ERHSQTS 3695
QY 1515 KSIDLTPGLLGGVVDLPESFPVVRMRQFVGCMRLQVDSHIDMADFIANNGTVPQCPAK 1574
DB 3696 YDIQELASALAHSVREIEHSVIGIRMR-----SALPVVPCQ 3730
QY 1575 KAVCDSTNCHNGGTGVNQMDAFSCCEPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLP1 1634
DB 3731 GQSCODQTCO-----ETVSLERVRVGSYSTARLSITLPHHIGRN-CSCNGTTLAR 3780
QY 1635 S-----QPVTLISLMTFRQADGVLLQALITGRSTITLQLREG--HYMLSVBEGT 1680
DB 3781 SGGSYVQYRPLEAQNMQIHFTLKLQFWALLM--FTNETASISLKLANGSHLEYHCPG- 3837
QY 1681 GLQASSIHLERPGANDGDHMAOLALGASGAPGHAISFPYGOQARAGNLPRLHGLHLS 1740
DB 3838 GFYGNLSSRRP--VNDGQMH--SMLEBRDTSVHLVDI---TDNASLVPRECGHAKTE 3890
QY 1741 -NITVGI--PGPAGYAGFRGCLQGVRVS-DTPGCVSLDPSHG--ESINVEQCGSLP 1794
DB 3891 ROLLAGGLVPSNPSNSVSLFPEGCLDAVVVNGERLELIGREKMEGRLEFMTALSQCCWPG 3950
QY 1795 DPDCDSNCPANVCSNMDWS-YGSCSDPGYTGDNQ--TNVCOLNPEHNSVCTRPSAP 1850
DB 3951 TACSQSCNLNGSGSPALSGSYLCCRPSPSGRNCBELGRENCTSAPOCBGCTCV---SSP 4007
QY 1851 HGVTCECPMYLGPYCESTR-----IDQPCPRGMGH-----PTCG 1885
DB 4008 EGTSCNCPHYTGBRCMEARCGSGHCLTTPHKKGDWQOQEPVITVALPLVIATVG 4067
QY 1886 -----PCNCDVSKGPPDCKNTG-----GECHCKENHRP 1915
DB 4068 LLLYCRRRKSHKPVYME---DPDLARSIGVDTQASPAIEPLDPLNTSSCN---MLNCP 4119
QY 1916 PGSPFCLCCPYTGLSLR---VCPBEDQCPG---KPGVIGROCDRCDNPAETTN 1967
DB 4120 BPSTISVPELVTFFGSSKORPMVCSVPRPLPRANVSHFG-----HEPLIKKTWS 4170
QY 1968 GCEVNDYSCPRALBAGTW---W-----PRTFGLPAAPCPKGS 2003
DB 4171 GBEELVYS-----GAAYMPRTYSKKKMEVPHPETMGTLPRPRRRHVIG--PAVMPDPTGL 4224
QY 2004 FG--TAVRHCDEHRGMLPEMLFN-----C--TSITP----- 2030
DB 4225 YGFPPLPLBENKAPLRPRYSNQNLEDLMPRPSPRPHILAPCLNEVYALISYHSQFR 4284
QY 2031 -----SELKGFARLORN-----BSGLDGRSQQOALLLRNATGHTAGYFG 2071
DB 4285 QGGGGPCLAGGTYGVSMKLSRAGPSYADCEVNGAPATGHSQRA-----PRNYEG 4335
QY 2072 SDV 2074
DB 4336 SDM 4338

```

## RESULT 8

T20721

hypothetical protein F25F2.2 - Caenorhabditis elegans

C/Spectrum: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 19-May-2000

C/Accession: T20721, T21343, T23842

R/Ainscough, R.

Submitted to the EMBL Data Library, August 1994

A/Reference number: Z19314

A/Accession: T20721

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;W13&gt;

A/Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone F10611

R/Ainscough, R.

Submitted to the EMBL Data Library, August 1994

A/Reference number: Z19410

A/Accession: T21343

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;W13&gt;

A/Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone F25F2

R/Suleston, J.

Submitted to the EMBL Data Library, June 1994

A/Reference number: Z19806

A/Accession: T23842

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;W13&gt;

A/Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone M88

C/Genetics:

A/Gene: CESP:F25F2.2

A/Map position: 3

A/Intons: 79/3; 167/3; 251/3; 328/3; 587/3; 625/3; 737/3; 1006/1; 1190/2; 1494/2; 1770/

C/Superfamily: rat MEGP1 protein, cadherin repeat homology; EGF homology; laminin G repe

F/761-3900/Domain: laminin G repeat homology &lt;IGR&gt;

Query Match

Best Local Similarity 21.6%; Score 1092; DB 2; Length 4307;

Matches 417; Conservative 248; Mismatches 641; Indels 628; Gaps 58;

QY 232 FSLDPTVGA-VTTAEE-----LDRRTKSTHVRV-----TAQDHGMRRSAL 272

DB 2410 FELAPTTSSNSTSKRAMPVALDRKTRGTHVSRVNLVYHDKRYQIPLVED--ATGRRAF 2467

QY 273 ATFTLVTDNDHDPVEQOEYKESLRNLEVGVEVLTVRATDGAAPPANILYRLBGS 332

DB 2468 STLTLSVIDIDINKPPFPVLPFYSTISBSAKGDTVMVVSATDD--ENDTIEYSLD-- 2523

QY 333 GSPSEVPEIDRSGVIRKRGVDRVEVESYOLTVEASQGDPPGRSTTAVPFISVEDD 392

DB 2524 -GSESQFSEVHRPQGTIVAKLKHAGVTLSTTKATDSANPPHATTVEV--NIASE 2580

QY 393 NDNAPOFSEKRYVVOVREDDVTPGAVLRYTASDRDKSAAVHVSIMSNGAQO----FY 448

DB 2581 SVKVRFSNSHLTFVMEADVGVNIGRQCMETIDE---IRFTVPSBSDSGLPSS 2637

QY 449 LDAQGALDVSPLDYETTKETTLAVRADGGRPPLSNVSGI--VTVQVLDINDNAPITV 506

DB 2638 VERSTGKTIIVKSLDRERKQKMTIRADAA--GVAHTVTVVDIGDVNDNAPAH 2692

QY 507 STPFOATVLESVPGLVYLHVQAIADAGDNARLEYRLAGVHDPPTINNCTGMSIVA 566

DB 2693 GYERTTIEDAAVGTSTVIFSAMDRDPSGRIRFSL--VEBNPYFDLNMESGMLTVAS 2750

QY 567 ELDRREVDFYSGVARDHGPALITASASVSVTVLVDVNDNPTFTQPEY----- 615

DB 2751 QLDREKIDHYKLIARATDGG--GFNTDLPFTIVSDVNDSPQFKEKEBFNIDLHPSTSP 2808

QY 616 ----- 615

```

Db      2809 ILHFSIKODDLSPNNVSQEFIPKNEGVWIDSNNDLLKRPETVENKQYQKLTARD 2868
Qy      616 -----TVRLNEDAAVGTSV-----VYTSAYDR 637
Db      2869 GVPETSTKYKYNLKSSKSDIRCPKANKVTLAENSKGTIVLGBESSLLGPNVTFGLSDN 2928
Qy      638 DAH-----SVITYQI-----647
Db      2929 DGNVFNVRNNGIYVKKESDQIDYEKQQLFHHRLTIQDNSEVCKELITIVLENLNDNRP 2988
Qy      648 -----TSGNTN-----RPSITSGSG-----GLVS 668
Db      2989 KLIKELKYSIDENLEPTSEDAQYITRIYAEADPDEIKFRWVDGGLFQJIDINGVYT 3048
Qy      669 LALPLDYKLERQYVLAVTSAGTROPATQIVNVNTDANTHRPVPQSHYTVANVERBPAG 728
Db      3049 VVKPLDSEILGFPNLTVAASDGEFEDKATILVTVDONNAPTFEKSTYSMTKMESESIG 3108
Qy      729 TVVVLISATDEDTGENARITYEMEDS-IFQF-RIDADTGAVTQAELDYEDQVSYTALIT 786
Db      3109 YELAHFRASGSD--QNETIETYLAKPSDVTSFVYNAQITILITAKPLDEFTLSALKLTIV 3166
Qy      787 ARDNQIPQKSDTFTYIELVNDVNDNAPOFLRDSYGSYEDVPPTSYLQISATDRDGL 846
Db      3167 AADSGVPLETRAQYVEISVWDEBNDNAKFEKEKYGVKENSKEKVLTVFALDVSSEH 3226
Qy      847 NGRVFPPTGGDGDGD-----FIVSTSGIVTRLRLRENAVOYVLAAYVDKMPA 901
Db      3227 FFAVSELEIVSEETTTDTPLVPALPAINS-NGDVLIQSSIDERYEIKKYNLVIYAKDGRPPL 3285
Qy      902 RTPMETVTVLVDNDNPVFEODEPVPVBEENSPIGLAVARVATPDDECTNAQIMQYIV 961
Db      3286 REBALLETIVBENBHAFTFDCCMTALVQEGEATLHLLKSVSLDPPKAGAPTYVI 3345
Qy      962 EKNIEPVFOLDIFSGELTALVDLYDRDEYVLT-----QATSAPLVSAVYHVRLL 1014
Db      3346 QGDGAKSFYVN-EKLELTAKKLEKYKODKYLTVAKOVKGTDCPL-----TIFIRQT 3400
Qy      1015 DENDNPPVLGNFELLFNNTVTRSSSPGGAGIYRVAHPDIDS---LTSFEGNELS 1071
Db      3401 SR-HAFTKPKMLQINTL--QNELPFGITIGRLASDEDEBDONGILRFLGVSIGS 3454
Qy      1072 -----LVLLNASTGEL-----KLSRALDNNRPLEAIVSVLSPGVASVTAQCAL 1115
Db      3455 PRAQVQESSTHLFRDPTGDIWSHDSITQGLHT-----FNTVYDSKNTSYSEV 3507
Qy      1116 RVTITTDKMLTSTLTLEDSPERPLSPULGLFIQAVATLATPPDHV-----VFN 1168
Db      3508 HTSISDNVIDIAVSRIRMSGVDEFRMKVKEFRRIISHHLNLDSSIQLSIVQAVPS 3567
Qy      1169 VORDTDAPEGHILNLSISVGPPGGGPPLPSBLOGL-----YANRSL 1216
Db      3568 TSESRSRNRSNMEDEILMTAORGLRG--YAKPDHYSRLKNDPQNMNDGQRMRYQOI 3625
Qy      1217 TAISAQVLPFPDNLICLRPPC-----ENYRCVSVLRFDSAPPIASSVLPFPIHVP 1269
Db      3626 TEMCT-----TGYCLRGKREVIELIDSWTKYSTDDFSFVSPPHSSA-----3669
Qy      1270 GGLRCCRPFGTADYCEIEVDLCYSRPGPHRC--RSREGGYTCLCRDGYTGEHCEVSA 1327
Db      3670 ---QCICPFGGKACEVETNOCAKSPCRQWOLCIPSVANSTYECVCPFGMEGDKCSVPS 3726
Qy      1328 RSGRCPCGVCKNGTCV---NLLVGGFKDCPSGDEKPYC-QVYTRSPPAHSFITFRL 1383
Db      3727 -----CQNDGKCLBAELSVGG-----DGFELISLSNEIETR-----3758
Qy      1384 RORFHETLALSPATKRDGLLYNGRFRNKHDFVALEVIOBOVOLTFNAGBSTTVSPV 1443
Db      3759 -----MELHIBLKTTHNGIIMWS---RGRDPTMLRLVNGTPEYHMANAGTGTGIVTS--3808
Qy      1444 PGVSDGQHHYVQAKTKYNKPLGQGLPGPBBQKAAVTVVGCCTGVALR-FGSVLGNY 1502

```

```

Db      3809 KTSVVDGQWHLIATIS-----RQRRTMTVDDDELDQEAFFSPISGTVINL 3852
Qy      1503 SCAAGTGGGSKKSIDLTLGPIILGGVPDLPSFPVRKQPVGCARMLOYDSNHT-----D 1557
Db      3853 HRYSQKLVLAGK-----VDDGELTGVSS-----ACFRTISVDGKRVKLTROG 3894
Qy      1558 MADFIANNNGVGCPCPA-KKAVCDNSNTCHNGGTGVNQMDA-FSCGCPILGFGKSCAQEMAN 1615
Db      3895 MCLF-----GAQPGCSALTSFPCNDLPQHAAGTCTISQGSNFRCEPCKPSKGNVCETDL--3948
Qy      1616 PQHFGSSLVAMHGLSLPISQPWTLSIMFRTRQADGVLLQAITRGRSTTYLQLRBHWML 1675
Db      3949 -----3948
Qy      1676 SVETGLOASSLRLEPGANDGDWHQALALGASGSPGHALLSPYGGQRAEAGNLGRLH 1735
Db      3949 -----3948
Qy      1736 GLHLNITVGGIPGPAGGVARGGLQGVVSDTPBEGVNSLDPSHGESINVEQCSLPD 1795
Db      3949 -----E 3949
Qy      1796 PCDNPNCPANSYCSNDWDSYSCSDPGYYGDN-----TNYCDLNPCEHGVCTRK 1846
Db      3950 FCASGCPCTGIGCIPIFYNDYLCCKPNGFTGKHCEARGFEDHETSCKNVCGTSGQCTSI 4009
Qy      1847 PS---AHGTYCEC 1857
Db      4010 PHSLSSDPFCNC 4023

```

## RESULT 9

T27110

Hypothetical protein Y52B11B.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T27110

R. Iennard, N. submitted to the EMBL Data Library, October 1998

A/Reference number: Z20311

A/Accession: T27110

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1439 &lt;NUL&gt;

A/Cross-references: EMBL:AL032638; PDB:CAA21627.1; GSDB:GN00019; CESP:Y52B11B.2

A/Experimental source: clone Y52B11B

C/Genetics:

A/Gene: CESP:Y52B11B.2

A/Map position: 1

A/Introns: 11/2; 103/1; 183/2; 299/1; 406/3; 502/1; 585/3; 884/2; 969/1; 1017/1; 1185/3;

Query Match 5.3%; Score 817; DB 2; Length 1439;

Best Local Similarity 28.5%; Pred. No. 1.5e-33;

Matches 319; Conservative 173; Mismatches 474; Indels 154; Gaps 52;

```

Qy      176 NTAPOFQPSYQATVPENQAGTPVASLRATDPDEGAGRLBYTMDALFDRSNQFSLD 235
Db      272 DNVPFTFRPLTYAQVREDIPINQTLIKTAVDQKDTGDSKRTIYSVD-----NNEFSIN 324
Qy      236 PVTAQVTTAEELD-----REKSTHVFRTYAQDHQP--RRSALATLTILVTDNDHPVF- 289
Db      325 S-NGSISAKVLDLADQLNERRHFVVRFTVADHGPVSLSSSAMHIHRTENTDSAVFL 383
Qy      290 EQGEKESLEENLAEVGYEVLTVRATDQDAPPNANILVRLLEGSGSPSE--VFETDPSG 347
Db      384 PTSQYTAFAVADAGQGFVIOIQARDAD--RDEVTYSFMDKNGRSYQKNLFSIDERTG 440
Qy      348 VIRTGFPV---DREAVES-YQITVEASDQGR---DPGRSTT-AAVFSVBDNDNAPOF 399
Db      441 LVKLRHGSADLAEANPILVTIVQDDGSCCVYPSKTHSYATLILIGIDVNNKKEF 500
Qy      400 SE-KRY--VVOVRDV--TPGAPVLRVYASDRDGSNAVHYST--MSGNARGOFTLDAQ 452

```

```

Db      501 PDCAKSIDIAKIMEGTYKTDPTIVKVEATDDSDANGDIVSLYYTOSSESRKAPVIDRQ 560
Qy      453 TGAADVVP---LDYETKEYLTVRAAOGGRPELNSVGLVYVQVLDINDNAPFVSTP 509
Db      561 TGYL-TPSRHVVDRETRPREVDYVKAITDRGRPLIGFCOF-SVEVVDINDNSPQERPS 618
Qy      510 PQATVLESVPLGYLVHQAIDADAGDNALREYL-----AGVGH-----DPEFTINGTG 560
Db      619 YETSVSRFEAVGTSVITVFAFNDNDAAHNAITVSLIEDITAGBEHQNDLDFELVVRASG 678
Qy      561 WISVAELDBREVDYFSGVEADHGT-ALTASASVSVTVLDVNDNPTF-TOPEYT-- 616
Db      679 EITLIPKIPKTKFI-FNVIAADDNGIPALOSSAQVTLAVLDKQAKAPMQSPCKKQ 737
Qy      617 VRLNEDAAVTSVVTASAVDR--DASVITYOITS-----GNTK-----NRPISOSG 663
Db      738 ITVDENVELAKVILRCRAVSSGDSRSDVYKLTASGGRNKAESFROPNK---ENG 793
Qy      664 GGLVSLAL--PLDYKEROVTLAVTASDGT-----QD-----TAQIVNV 702
Db      794 NENVEVIMEGLDYEQVNNVTLTLTATVGPVPIQEKVCFDFQDMTSRVASTKEFVEV 853
Qy      703 TDANTRPVPOSHYVNVNVE-----DRPAGTVVLISATDEPT-GENARITTFM-- 751
Db      854 RVNDVVPFQVVDLEFGTIDEEMTPNEHLEKTKGKIVYVKAIDTSDGPONEVHYRIYG 913
Qy      752 ---EDSIPQRIADAGVTTQAEIDYEDQVSYTLAITARDNGI-----PQSDTTY 800
Db      914 EANGETTKFRIDELTGEIFPNEKFPREKIDMTITLVVEASDRSVSLPGANGPNK-DNK 972
Qy      801 LELVNDVNDNAQPLRDSYQSVYEDVPEFTSVLQISATDRDGLNGRVFTYPOGDDG 860
Db      973 VOIVINDVNDNAPFEQKTYIGVKESEGEHDVITIKADLDKHSNLR--YHLIGAGG 1030
Qy      861 DGGFVYESTSGIVRTLRDLRENNVAVYLRAYVNDGMPARTPMVYTVVLDVNDNPPV 920
Db      1031 RLPFGVRTSGTIFVKEPLDFEASDQYHLVLASD--GRHNAT--NVYHIEIDVNNNAQ 1087
Qy      921 FEODEF-DVFEVENSPIGLAVAVTATDPEG-TNAQIMQIVEGNIPEVFOLDIFSGEL 978
Db      1088 FPGQKATVIEEDVDIPIKVLFWVHATDADQDEKSRIVYRLGQADAVFRIGKSGTI 1147
Qy      979 TALVDLDYDR--PEYVLYIQATSNP--LYSRATVHYRLDRNDPVLGNFEILFNN 1032
Db      1148 ELVKALDRPPAGVPMNFVQAIIDDGNGLVGVAOVVNDJINNSPIFP--ERLFGY 1205
Qy      1033 YVTNRSSSPGGA-IGRVPAHDIDISLTYSEFG-----NELSVLNASGELK 1084
Db      1206 IEBNRPISDGVYFMDVQARDFDPTTENANIEYIVRNKLINGESVFRIDQNTKIRA 1265
Qy      1085 SPALDN--NRPLRAIMSVLSD-GVHSV--AQCALRVTLITDEMILTHSITL--RLSDM 1136
Db      1266 MSLDIBISERFPILEVANDGVSRSGFANVTIKVTDMDNNAFPFEKIRIEGSEVET 1325
Qy      1137 SPERFLSPILGFIQAVATLA--TPPDHVVENVQRTD 1174
Db      1326 AP-----IGAAMSFSAFADDEBAKDVFTQLSESD 1358

RESULT 10
T00043
BH-protocadherin-a - mouse
C/Species: Mus musculus (house mouse)
C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C/Accession: T00043
R/Yoshida, K.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z14075
A/Accession: T00043
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1069 <YOS>
A/Cross-references: EMBL:AB06758; NID:d1227200; PIDN:BA32597.1; PID:d1033562

```

```

C/Genetics:
A/Gene: Pcdh7
A/Map position: 5C3-D
Query Match 4.9%; Score 768.5; DB 2; Length 1069;
Best Local Similarity 28.1%; Pred. No. 2.8e-31;
Matches 224; Conservative 128; Mismatches 339; Indels 107; Gaps 17;

Qy      328 LLEGSG-----GSSVEFEIDPRSGVIRT--RGVDRREVBESYOLTVASDQGRD----- 375
Db      57 IYTGSEVTFPSLESSEYKIDNLTGELISTERRIDREKLPQCCMTFDBNECFLDPEVSV 116
Qy      376 PGPSTTAVF---LVEDDNDNAPQFSEKRYVQVREDVTEGAVLVTAASDRDGSNA 432
Db      117 IGPQSWDLFEGRIVLDINDNPTFPSPVLTLLVENRPVGTLYLPLTADRPGNG 176
Qy      433 VVHSIM-----SGNARQTYLDAQTA---- 455
Db      177 IERYELQEPGGGGSGGRRLGPADSAPYPOGGNSASGGSGSKRLDAPEGGGGTS 236
Qy      456 -----LDVSPLDYETTKETLRYRAQDGRPLSNVSGLYT 492
Db      237 FSGRSVELQVADTPDSEKQPLIVKCALDBEQDSYELTLRVGDGDPFSS--QALIR 295
Qy      493 VQVLDINDNAPFVSTPQATVLESVPLGYLVHQAIDADAGDNALREYRLAGVGHDP 552
Db      296 VLTVDNNSPREFKSVYEAADLAENSAPGTPILQRLATDLDVGVNGIEYVGAATESVR 355
Qy      553 --FTTNGTGMISVAELDBREVDYFSGVEARDHGTALTASASVSVTVLDVNDNPT- 609
Db      356 RLRLIDETSGMLSVLRIIDREVNQARFTMARBGQPKTKATVNVINDENNVPSI 415
Qy      610 -----FTPEYTVRLNEDAAVGSVTVASVND--AHSVITYQISGTRNRFITS 660
Db      416 EIRTKGRPLADGANVADVLVTPILAVQVSRDQSENGVYTTV-GDVPQLKAS 474
Qy      661 QSGGG-----LVSLALPLDYKLEQY--VLAVTASDGTQDTAQIVNVTTDANTHRV 711
Db      475 DTEGQNKKKYFLHTSAPLDYETTFEYFNVIVAVDSGPSLSNNSLVKVGDTINDNPPV 534
Qy      712 FQSHYTVNVEDRAGTVVLISATDEPTGENARITFEMDSI--PQRIADTCAVTTQ 770
Db      535 FQGSVEYVFPENNIPGRVATVLTATDADSGNMEIASLSWSVGTALIDPSSGILVN 594
Qy      771 AELDYEDQVSYTLAITARDNGIPQSDTTYLEILVNDVNDNAPQLRDSYQSVYEDVP 830
Db      595 TVLDBEQDRIFKKNADKQIPVLQSGSTVIYQVADKNDPKRMQDVFTTYVENIQP 654
Qy      831 FTSVLQISATDRDGLNGRVFTYPOGDDGDPFVSTSGIVRTLRLDRNNVAVYLR 890
Db      655 NSPVMVTVMDADKGRNAMESLYI--BENSNIFSIENDGTIVSTWGFDRHQTYTFR 711
Qy      891 AYAVDKGMPARTPMHVTVVTVLDVNDNPP--VFQDEDDVYBENSPIGLAVARTADP 948
Db      712 VKAVDGDPSPSARTAVSLFVMDENDNAPVTLPPNISYTLTPSSNVRTVAVALADDS 771
Qy      949 DEGTAAQIMQIVEGNIEVEFQLDIFSGELTALVDLYEDREPEYLVYQATSAPLVSRAT 1008
Db      772 DDGINADLNTSGIVGNPFKLPFIDISTSGVSLVNGTLQKHGIAHRLVYQVNDSGPQST 831
Qy      1009 ---VHRLDRNDNPVL 1023
Db      832 TTLVHVFNESVSNATVI 849

```

```

RESULT 11
T00041
BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human
C/Species: Homo sapiens (man)
C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 01-Dec-2000
C/Accession: T00041; T00040
R/Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.; Sugano, S.
Genomics 49, 458-461, 1998

```





```

Db      1002 IGVSPFVIVNLVFLWTLTHKQVRSSESVLPDSSRLDNKSMALGAILFLGLTWARG 1061
Qy      2580 LLSVNSDTLLFHYLFATCNCTIOGPIFLSYVLSKEVRKAL-----KLACSRKSPDPA 2633
Db      1062 LFLINKESVMAVYLFETTFNAFQGVFLFPCALQKVKHKEYSKCLRHSTCCIRNP---PG 1118
Qy      2634 LTTKSTLTSSYNCSPSPYADG---RLYQPYGDS-----AGSLHST----- 2669
Db      1119 GAGSLKTSAMRSNARYTGTQSRIRRMNDIVRKQTESFPMAGDINSTTLTARGTMGNH 1178
Qy      2670 -----SRSGSGQSPYIFP---LLREESALNPGQGPFGADPGSLFLESGDQDQHPDPTD 2719
Db      1179 LLTNPVLQPRGTS-----PYNTLILASVGFNP--SSPFPVNSPGS--YREPKHPLGSGREAC 1231
Qy      2720 SDSDSLSLDDQSGSYASTHSSDSSEEBEERBEAAPGEGQWDSLPGAEKRLP----- 2773
Db      1232 GMDTFLPLNGFNNSY--SLRSGD-----PPP-----GDGAPPPGRGNIA 1269
Qy      2774 -----LHSTPKDG-----GPGPKAPWPGDPFGTTAKESGNGAPPEERLRENG 2815
Db      1270 DAAPAEKMIISLVHNNLRGGSSGAKGPPPEPPVPVPVPGSGGEBAAG----- 1318
Qy      2816 DALSRBSGLGPLGSSAQP---HKGLKKKCLPTISEKSLRLPLEQCTGSSRGSSAS 2871
Db      1319 -----PGADRABIEILLYKALEBPLLP--RAQSVLYQSDIDE-----SESCYAE 1360
Qy      2872 EGSRGGPPRPAPPROSL 2888
Db      1361 DGATSRPLSSPPGRDSL 1377

```

## RESULT 13

```

T18413
latrophilin-1, brain-specific - bovine
N/Alternate names: alpha-latrotoxin receptor, calcium-independent
C/Species: Bos primigenius taurus (cattle)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
A/Accession: T18413
R/Matsuhashita, H.; Ieliano, V. G.; Ushkaryov, Y. A.
FEBS Lett. 443, 348-352, 1999
A/Title: The latrophilin family: multiply applied G protein-coupled receptors with differ
A/Reference number: Z18869; M0UD:99148828; PMID:10023961
A/Accession: T18413
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1472 <MAT>
A/Cross-references: EMBL:AF11098; NID:g4185803; PID:g4185804; PIDN:AAD09192.1
A/Experimental source: Brain
C/Genetics:
A/Gene: LPH1
C/Superfamily: alpha-latrotoxin receptor, calcium-independent
C/Keywords: G protein-coupled receptor

```

```

Query Match      4.7%; Score 729.5; DB 2; Length 1472;
Best Local Similarity 25.9%; Pred. No. 4.5e-29;
Matches 269; Conservative 139; Mismatches 378; Indels 251; Gaps 42;

```

```

Qy      1969 CEVVYDSCPRAIEAGIMWPTREGLPAAPCPKSGPETAARHCDERHG--WLP--PMLFNC 2025
Db      480 CE-----PRRVR--VQMPATQGMVVERPCPKTRGIAFOCLPALGLMNPSPDLSNC 532
Qy      2026 TSITSELKGFARLQNRSGSLDSRQQLALLRNATQHT-----AGYGSQVAVAYQL 2080
Db      533 TSPVNVGV---AQIKSGENA-----ANTASLARHTRGSIVAGDVSSVKKMEQL 580
Qy      2081 -----ATRLAHSTQRGSLATQD-----VHTENILARGSALDLTANK 2121
Db      581 LDILDAQLQALRPIERSAGKNNVKKHKBRTCKDYIKAVETVDNLR--PEAL----- 633
Qy      2122 RHMELIQOTGGG--TAMLQHYEAYASALAKNNHTLSPFTIYTPVIVISVRLD--KGN 2178
Db      634 ESMQDMNATQAHATATMLDVLLEBGAFLADNVRFP--AATFLAKQNVVLEVIVANREQ 691

```

```

Qy      2179 FAGALTPRYEALRGQPPDLKTTVILPESVFELETPPVVRPAGPGEAOBPEELARQRHP 2238
Db      692 VQELVFPQ-----EYPS--ENSIQLSANTIKN----- 717
Qy      2239 ELISQEBANASVIYITTLAGLLPHNYD-----PDKRSIVRKPRPIINTPVVSI 2285
Db      718 --SRNGVAVVPIILYNNLGLPLSTENATVKLAGESGPGGASL-----VNSQVILAA 769
Qy      2286 SVHDEBELPRALDVPYVQPLLTETERTKPICYFMNHSILVSGTGSGANSGCVFEN 2345
Db      770 STNKESKRY--FLMDPVLFTVAHLAKHFNANCSFMMVSEB--SMLGTWSTQCGLLVSN 826
Qy      2346 ESHVSCQCNHMTSPFVMDVSRRENGEI--PLKTLTYVALGVTLAALLTFEFLTLRI 2403
Db      827 KHTTLCASHLINFVLAHREIYQGRINELLSTVWGVILSVLCALICSTCPLRG 886
Qy      2404 LRSNQGIRNITAAQLAQLVFLGINQADLPACTYAIALLHFLYICTSMALLBLAH 2463
Db      887 LQTDNNTIHKNCIMFLABELFLVIGIDTQYEIACPIFAGLIHYFLAASMLCLEGVH 946
Qy      2464 LYRALTEVRDVNTGMRFYVNLGCVPAFTGLAVGLDEGYNDFCMLSIYDTLIWSF 2523
Db      947 LYLILVEFESERYSTRKTYTLGAGYCPFALVGIATAIDRSYGEKACMLRYDNTFINSF 1006
Qy      2524 AGPVAFAVMS--VFLYIL---AARASCAORQGEKGPVSGLOPSFAVLLLSATWLLA 2579
Db      1007 IGVSPFVIVNLVFLWTLTHKQVRSSESVLPDSSRLDNKSMALGAILFLGLTWARG 1066
Qy      2580 LLSVNSDTLLFHYLFATCNCTIOGPIFLSYVLSKEVRKAL-----KLACSRKSPDPA 2633
Db      1067 LFLINKESVMAVYLFETTFNAFQGVFLFPCALQKVKHKEYSKCLRHSTCCIRNP---PG 1123
Qy      2634 LTTKSTLTSSYNCSPSPYADG---RLYQPYGDS-----AGSLHST----- 2669
Db      1124 GAGSLKTSAMRSNARYTGTQSRIRRMNDIVRKQTESFPMAGDINSTTLTARGTMGNH 1183
Qy      2670 -----SRSGSGQSPYIFP---LLREESALNPGQGPFGADPGSLFLESGDQDQHPDPTD 2719
Db      1184 LLTNPVLQPRGTS-----PYNTLILASVGFNP--SSPFPVNSPGS--YREPKHPLGSGREAC 1236
Qy      2720 SDSDSLSLDDQSGSYASTHSSDSSEEBEERBEAAPGEGQWDSLPGAEKRLP----- 2773
Db      1237 GMDTFLPLNGFNNSY--SLRSGD-----PPP-----GDGAPPPGRGNIA 1274
Qy      2774 -----LHSTPKDG-----GPGPKAPWPGDPFGTTAKESGNGAPPEERLRENG 2815
Db      1275 DAAPAEKMIISLVHNNLRGGSSGAKGPPPEPPVPVPVPGSGGEBAAG----- 1323
Qy      2816 DALSRBSGLGPLGSSAQP---HKGLKKKCLPTISEKSLRLPLEQCTGSSRGSSAS 2871
Db      1324 -----PGADRABIEILLYKALEBPLLP--RAQSVLYQSDIDE-----SESCYAE 1365
Qy      2872 EGSRGGPPRPAPPROSL 2888
Db      1366 DGATSRPLSSPPGRDSL 1382

```

## RESULT 14

```

T17145
CLAB protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
A/Accession: T17145
R/Sugita, S.; Ichtchenko, K.; Khvorochey, M.; Sudhof, T. C.
submitted to the EMBL Data Library, July 1998
A/Description: CL family.
A/Reference number: Z18712
A/Accession: T17145
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1510 <SIG>
A/Cross-references: EMBL:AF081145; NID:g3695116; PID:g3695117; PIDN:AA62651.1

```

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 4.7%; Score 728; DB 2; Length 1510;  
 Best Local Similarity 26.0%; Pred. No. 5.6e-29;  
 Matches 278; Conservative 133; Mismatches 387; Indels 270; Gaps 42;

QY 1969 CEVNTDSCRAIEAGIWMPTRTFGLPAAAPCPKSGFGTAVRHCDEHRG-WLP-PNLFNC 2025  
 DB 475 CE-----PREVR-VQMPATQGMVERPCPKGTGASFOCLPALGLMNRGPDLSNC 527  
 QY 2026 TSITSEIKGFAERLORNESGLDSGSQLALLRNATQHT-----AGFSGDYKAYQL 2080  
 DB 528 TSPWNVQV---AQIKSGENA-----ANISLARHTGSIYAGDVSSVTKMEQL 575  
 QY 2081 -----ATRLAHSTORFGLSATOD-----VHFTENLRVGSALIDTANK 2121  
 DB 576 LDILDAOLALRPIERESAGKNVNMKHKERTCKOYIKAVETVDNLK-PEAL----- 628  
 QY 2122 RHWELIQOTE--GGTAMLOHTYEAVASALAQNKRHTTSLPFTITPNIIVISVRLDKGNF 2179  
 DB 629 ESMKQMNATEQVHTATMLDVLEEGAFLLADNVR-----EPARF 667  
 QY 2180 AGAKLPREYALRGEPDLETTVILPESVFRETPPVVRPAGPGEAOE---PEELARQOR 2236  
 DB 668 LAAK-----QNVLEVTVLSTE-----GVQELVFPQEVYA--SES 700  
 QY 2237 HPELS-----QGEAVASVIYRTLAGLPHNY-----DPDKSLRVPKR 2275  
 DB 701 SIQLSANTIKONSRNGVAVKVPILYNNMLFLSTENATVYKLAGAGTGGPGASL----- 755  
 QY 2276 PIINTPVVISVHDBELLPRALDKPVTYQFRLSTEERTKPICVFNNHSILVSGTGWS 2335  
 DB 756 -VNSQVLAASINKSSRV--FLMDPVITFYVAHLEAKHFNANCSFNNYSER-SMLGWS 811  
 QY 2336 ARGEVFRNESHVSCQNMHTSPAVLMDVSRNGEI--LPLKTLTYVALGVTALALL 2393  
 DB 812 TQGCGLVBSNKTHTTACSHLTNFAVMAHREIYQGRINELLSVTWGIYSLVCLAI 871  
 QY 2394 TFFPLTLRLIRSNQHGIRNLTAAAGLAQVLGLINQADLPACTVIALILHFLYICT 2453  
 DB 872 CISTCFPLRGLOTDNRNTHKNCINLFLABELFLVIGIDKTYEVAACPFIAGLHFFFLAA 931  
 QY 2454 FSWALLLEALHYRALTEVDVNTGPMRFYTMKGVPAPITGLAVGLDEGYNPDCWL 2513  
 DB 932 FSWMLLEGVHLYLLVEVESEYSRTKYIYLGICFPALVGIAAIDRSGTGAACWL 991  
 QY 2514 SIYDTLIWSFAGPVAFAVMS-VFLYIL--AARASCAORGFKKGPVSGLOPSPAVL 2569  
 DB 992 RVDNTPINSFIPGVSFVIVNLVFLMVTILHKMIRSSSVLKPDSSRLDNIKSMALGAILL 1051  
 QY 2570 LLLSATWLLALSVNSDTLLFHYLPATCNCIOGPFIFLSVYLSKVRKAL-----KLA 2623  
 DB 1052 FLGLGTMAFGLPIFKESVMAVLFITTPAPQGVIFIVPHCALQKKVKEYKCLRHGYC 1111  
 QY 2624 CSRKSPDPAALTTKSTLTSSYNCPSPYADRLYQPYGDSAGLSHTSNGSQPSYIFPL 2683  
 DB 1112 CTRSP---PGAHGSLIKTSAMRSNTRYTGTVPGGQ--RHIIHVSJLGRGR-SALP-- 1162  
 QY 2684 LMBESALNCGGPGGLGDP-----GSLFLBQDDQHDHDTSD 2721  
 DB 1163 ---BSQKDPG-QGSGPBDLITGLCPSRIRRMWMDTVRKQTESPMAG-----DIN 1209  
 QY 2722 SDLSLEDDQSGSVASTHSDSEEBEEREAAP-----GEOGMDSL- 2765  
 DB 1210 STPLINRGTMGNHLLTN-----PVLQGRGTSYNTLLAESVGFNPS 1252  
 QY 2766 -----GPAERLPIASTPMDGPGPKAWPDPDFTGAKSS-----GNGAPE-ERLKEN 2814  
 DB 1253 PEFVNSPGSYPREPKPLGREGACMDTLPPLNGFNNSYSLRSGDPPDPDGGGEBPRGNLT 1312  
 QY 2815 GDALRE-----GSLGPLPGSSAOPKHGILKKKC-LPTISSEKSL--- 2853  
 DB 1313 ADAAAFKMIITSELVHNNLRGASGAKGPPPPPPVPVPVGVSEDAAGPBGADRAIELL 1372

QY 2854 -----LRLPLEQCT-----GSSRGSSASBGRSGCPPPRPPRQSL 2888  
 DB 1373 YKALEEPLILRPAQSVLYQSDLBSESCTADGARSPLSSPPGDSL 1420

RESULT 15  
 T17156  
 CLUBB protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T17156  
 R:Sugita, S.; Ichchenko, K.; Khavochey, M.; Sudhof, T. C.  
 submitted to the EMBL Data Library, July 1998  
 A:Description: CL family.  
 A:Reference number: Z18712  
 A:Accession: T17156  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1515 <8UG>  
 A:Cross-references: EMBL:AF081147; NID:G3695120; PID:G3695121; PIDN:AA62653.1

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 4.7%; Score 728; DB 2; Length 1515;  
 Best Local Similarity 26.0%; Pred. No. 5.6e-29;  
 Matches 278; Conservative 133; Mismatches 387; Indels 270; Gaps 42;

QY 1969 CEVNTDSCRAIEAGIWMPTRTFGLPAAAPCPKSGFGTAVRHCDEHRG-WLP-PNLFNC 2025  
 DB 480 CE-----PREVR-VQMPATQGMVERPCPKGTGASFOCLPALGLMNRGPDLSNC 532  
 QY 2026 TSITSEIKGFAERLORNESGLDSGSQLALLRNATQHT-----AGFSGDYKAYQL 2080  
 DB 533 TSPWNVQV---AQIKSGENA-----ANISLARHTGSIYAGDVSSVTKMEQL 580  
 QY 2081 -----ATRLAHSTORFGLSATOD-----VHFTENLRVGSALIDTANK 2121  
 DB 581 LDILDAOLALRPIERESAGKNVNMKHKERTCKOYIKAVETVDNLK-PEAL----- 633  
 QY 2122 RHWELIQOTE--GGTAMLOHTYEAVASALAQNKRHTTSLPFTITPNIIVISVRLDKGNF 2179  
 DB 634 ESMKQMNATEQVHTATMLDVLEEGAFLLADNVR-----EPARF 672  
 QY 2180 AGAKLPREYALRGEPDLETTVILPESVFRETPPVVRPAGPGEAOE---PEELARQOR 2236  
 DB 673 LAAK-----QNVLEVTVLSTE-----GVQELVFPQEVYA--SES 705  
 QY 2237 HPELS-----QGEAVASVIYRTLAGLPHNY-----DPDKSLRVPKR 2275  
 DB 706 SIQLSANTIKONSRNGVAVKVPILYNNMLFLSTENATVYKLAGAGTGGPGASL----- 760  
 QY 2276 PIINTPVVISVHDBELLPRALDKPVTYQFRLSTEERTKPICVFNNHSILVSGTGWS 2335  
 DB 761 -VNSQVLAASINKSSRV--FLMDPVITFYVAHLEAKHFNANCSFNNYSER-SMLGWS 816  
 QY 2336 ARGEVFRNESHVSCQNMHTSPAVLMDVSRNGEI--LPLKTLTYVALGVTALALL 2393  
 DB 817 TQGCGLVBSNKTHTTACSHLTNFAVMAHREIYQGRINELLSVTWGIYSLVCLAI 876  
 QY 2394 TFFPLTLRLIRSNQHGIRNLTAAAGLAQVLGLINQADLPACTVIALILHFLYICT 2453  
 DB 877 CISTCFPLRGLOTDNRNTHKNCINLFLABELFLVIGIDKTYEVAACPFIAGLHFFFLAA 936  
 QY 2454 FSWALLLEALHYRALTEVDVNTGPMRFYTMKGVPAPITGLAVGLDEGYNPDCWL 2513  
 DB 937 FSWMLLEGVHLYLLVEVESEYSRTKYIYLGICFPALVGIAAIDRSGTGAACWL 996  
 QY 2514 SIYDTLIWSFAGPVAFAVMS-VFLYIL--AARASCAORGFKKGPVSGLOPSPAVL 2569  
 DB 997 RVDNTPINSFIPGVSFVIVNLVFLMVTILHKMIRSSSVLKPDSSRLDNIKSMALGAILL 1056  
 QY 2570 LLLSATWLLALSVNSDTLLFHYLPATCNCIOGPFIFLSVYLSKVRKAL-----KLA 2623

```

Db      1057 FLGLTWAFGLLPINKESVMAYLFTTNAFCGVFIVFHICALQKKVHKRYSKCLRHASYC 1116
QY      2624 CSKRPSFDPALTTKSTLTSSYNCPYADGRLYOPYGDSAGSLHSTSRSGSQPSYI PFL 2683
Db      1117 CIRSP---PGAGAGSLKTSAMRSVTRYTGTVGOG---RHITHOVSLGPRGR-SALP-- 1167
QY      2684 LREESALNPGQGPGLGDP-----GSLFLEGQDQOHDPDTDS 2721
Db      1168 ---BQKDPG-GQSGPDPDLTFGLCPGRIRRMNDTVRKQTESSFMAG-----DIN 1214
QY      2722 SDSLLEDDQGSYASTHSDSSEEEEEEAAFP-----GEGQMDSL- 2765
Db      1215 STPTLRNGTGNHLLTN-----PVLQPRGTSPTVTLIABSVGNPSS 1257
QY      2766 ----GPGARLPLHSTPKDGGPGKAPWPGDGTAKES-----GNCAPE-ERLREN 2814
Db      1258 PVTNSPGSYREPKHPAGREACGMDTLPLNGVFNNSYSLRSGDPPPGDGPPEPRGRNL 1317
QY      2815 GDALSRE-----GSLGPLGSSAQPHKGLKKC-LPTISEKSSL--- 2853
Db      1318 ADAAAFKMTISLHVNNLRGASGAGAKPPPPVPPVPGVSDDEAGGPGGADRAABIELL 1377
QY      2854 -----LRLPLECT-----GSSRGSASBSRGSGPPPPROSL 2888
Db      1378 YKALEEPLLLPRAQSVLYQSDLDSEBECTAEDGATSRPLSSPPGRDSL 1425

```

Search completed: February 11, 2004, 15:57:55  
 Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:51:34 ; Search time 25 seconds  
(without alignments)

5498.362 Million cell updates/sec

Title: US-09-916-849a-3

Percent score: 15545  
Sequence: 1 MRSFATGVLPPTPPPLLL.....ACTVDESSGSEFLPFNFLLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15545	100.0	2923	1	CLR2_HUMAN
2	14657.5	94.3	2920	1	CLR2_MOUSE
3	11097	71.4	2144	1	CLR2_RAT
4	8974.5	57.7	3034	1	CLR1_MOUSE
5	8754	56.3	3014	1	CLR1_HUMAN
6	7833.5	50.4	3312	1	CLR3_HUMAN
7	7780.5	50.1	3313	1	CLR3_RAT
8	7732.5	48.7	3301	1	CLR3_MOUSE
9	4817.5	31.0	3579	1	STRN_DROME
10	1534.5	9.9	5147	1	FATP_DROME
11	1516.5	9.8	4590	1	FATP_HUMAN
12	1477.5	9.5	4705	1	FAT2_DROME
13	1419.5	9.1	3097	1	CADN_DROME
14	1320.5	8.5	4349	1	FAT2_HUMAN
15	1275	8.2	4351	1	FAT2_RAT
16	1262.5	8.1	3398	1	PC16_HUMAN
17	1218.5	7.8	3317	1	CADN_RAT
18	1217.5	7.8	3354	1	CADN_HUMAN
19	1214	7.8	3354	1	CADN_MOUSE
20	1207	7.8	3503	1	DS_DROME
21	985	6.3	2215	1	CD82_DROME
22	905	5.8	1967	1	CD87_DROME
23	858	5.5	1507	1	CAD8_DROME
24	853	5.5	1955	1	PC15_HUMAN
25	846	5.4	1026	1	PC15_HUMAN
26	833.5	5.4	1943	1	PC15_MOUSE
27	764.5	4.9	1069	1	PC17_HUMAN
28	710.5	4.6	1584	1	PC17_HUMAN
29	687.5	4.4	3343	1	YOC2_HUMAN
30	678.5	4.4	1007	1	CHC2_MOUSE
31	668.5	4.3	2240	1	CHC2_HUMAN
32	667.5	4.3	793	1	CHC2_DROME
33	667	4.3	936	1	CD87_HUMAN

34	662.	4.3	798	1	CD8E_HUMAN	Q9559 homo sapien
35	654	4.2	795	1	CD8C_HUMAN	Q9551 homo sapien
36	653	4.2	1522	1	BA13_HUMAN	O60242 homo sapien
37	649.5	4.2	950	1	CD43_HUMAN	Q9558 homo sapien
38	648.5	4.2	1180	1	PC12_MOUSE	O55134 mus musculu
39	647.5	4.2	950	1	CD41_HUMAN	Q95510 homo sapien
40	644.5	4.1	798	1	CD8D_HUMAN	Q95511 homo sapien
41	642	4.1	949	1	CD8E_HUMAN	Q95510 homo sapien
42	632	4.1	801	1	CD8B_HUMAN	Q95510 homo sapien
43	630	4.1	947	1	CD8A_HUMAN	Q95510 homo sapien
44	628.5	4.0	795	1	CD84_HUMAN	Q95510 homo sapien
45	628.5	4.0	797	1	CD83_RAT	Q63418 rattus norv

## ALIGNMENTS

RESULT 1  
ID CLR2\_HUMAN STANDARD; PRT; 2923 AA.  
AC O9HCT4; Q92566;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cadherin BGF lag seven-pass G-type receptor 2 precursor (Epidermal growth factor-like 2) (Multiple epidermal growth factor-like domains 3) (Flamingo 1).  
DE CHER2 OR CDH10 OR EGF2 OR MEG3 OR KIAA279.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363102; PubMed=10907856;  
RA Vincent J.B., Skaug J., Scherer S.W.,  
RT "The human homologue of flamingo, EGF2, encodes a brain-expressed large cadherin-like protein with epidermal growth factor-like domains, and maps to chromosome 1p33.3-p21.1.",  
RT DNA Res. 7:233-235(2000).  
RL [2]  
RM SEQUENCE OF 516-2923 FROM N.A.  
RN TISSUE=Brain;  
RC MEDLINE=97191544; PubMed=9039502;  
RX Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y.,  
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.,  
RT "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by RT analysis of cDNA clones from cell line KG-1 and brain.",  
RT DNA Res. 3:321-329(1996).  
RL [1]  
CC - FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
CC - SUBCELLULAR LOCATION: Integral membrane protein.  
CC - TISSUE SPECIFICITY: Highest expression in brain and testis.  
CC - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC - SIMILARITY: Contains 9 cadherin domains.  
CC - SIMILARITY: Contains 8 BGF-like domains.  
CC - SIMILARITY: Contains 2 laminin G-like domains.  
CC - SIMILARITY: Contains 1 GFS domain.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: AF234887; AAC00080.1; -  
CC EMBL: D87469; BA13407.1; -  
CC HSSP: P15116; INCU.  
CC Genew: HGNC:3231; CHER2.  
CC MIM: 604265; -

DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0004930; F:G-protein coupled receptor activity; NAS.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR000832; GPCR secretin.  
 DR InterPro; IPR001879; hormn. receptor.  
 DR InterPro; IPR002049; laminin EGF.  
 DR InterPro; IPR001791; laminin G.  
 DR InterPro; IPR00203; PLO\_cys-rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00053; laminin\_EGF; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00282; Hormr; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 7.  
 DR PROSITE; PS00268; CADHERIN\_2; 9.  
 DR PROSITE; PS00221; GPS; 1.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS00651; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS00652; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS00653; LAM\_G\_DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal.  
 FT CHAIN 1 31  
 FT SIGNAL 32 2923  
 FT DOMAIN 32 2380  
 FT TRANSSEM 2381 2401  
 FT TRANSSEM 2402 2416  
 FT TRANSSEM 2417 2437  
 FT TRANSSEM 2438 2438  
 FT TRANSSEM 2439 2459  
 FT TRANSSEM 2460 2480  
 FT TRANSSEM 2481 2501  
 FT TRANSSEM 2502 2519  
 FT TRANSSEM 2520 2540  
 FT TRANSSEM 2541 2560  
 FT TRANSSEM 2561 2581  
 FT TRANSSEM 2582 2591  
 FT TRANSSEM 2592 2612  
 FT TRANSSEM 2613 2923  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 505  
 FT DOMAIN 506 610  
 FT DOMAIN 611 712  
 FT DOMAIN 713 815  
 FT DOMAIN 816 921  
 FT DOMAIN 922 1023  
 FT DOMAIN 1028 1146  
 FT DOMAIN 1228 1286  
 FT DOMAIN 1288 1324  
 FT DOMAIN 1328 1366  
 FT DOMAIN 1367 1571

FT DOMAIN 1574 1610 EGF-LIKE 4, CALCIUM-BINDING.  
 FT DOMAIN 1614 1791 LAMININ G-LIKE 2.  
 FT DOMAIN 1793 1828 EGF-LIKE 5, CALCIUM-BINDING.  
 FT DOMAIN 1829 1867 EGF-LIKE 6, CALCIUM-BINDING.  
 FT DOMAIN 1883 1922 EGF-LIKE 7, CALCIUM-BINDING.  
 FT DOMAIN 1923 1955 EGF-LIKE 8, CALCIUM-BINDING.  
 FT DOMAIN 2316 2368 GPS.  
 FT DOMAIN 2743 2752 POLY-GLU.  
 FT DISULFID 1232 1243 BY SIMILARITY.  
 FT DISULFID 1237 1274 BY SIMILARITY.  
 FT DISULFID 1276 1285 BY SIMILARITY.  
 FT DISULFID 1292 1303 BY SIMILARITY.  
 FT DISULFID 1297 1312 BY SIMILARITY.  
 FT DISULFID 1314 1323 BY SIMILARITY.  
 FT DISULFID 1332 1343 BY SIMILARITY.  
 FT DISULFID 1337 1353 BY SIMILARITY.  
 FT DISULFID 1355 1365 BY SIMILARITY.  
 FT DISULFID 1378 1389 BY SIMILARITY.  
 FT DISULFID 1578 1589 BY SIMILARITY.  
 FT DISULFID 1583 1598 BY SIMILARITY.  
 FT DISULFID 1600 1609 BY SIMILARITY.  
 FT DISULFID 1797 1808 BY SIMILARITY.  
 FT DISULFID 1802 1817 BY SIMILARITY.  
 FT DISULFID 1819 1828 BY SIMILARITY.  
 FT DISULFID 1832 1843 BY SIMILARITY.  
 FT DISULFID 1837 1855 BY SIMILARITY.  
 FT DISULFID 1857 1866 BY SIMILARITY.  
 FT DISULFID 1887 1899 BY SIMILARITY.  
 FT DISULFID 1889 1906 BY SIMILARITY.  
 FT DISULFID 1908 1921 BY SIMILARITY.  
 FT DISULFID 1924 1936 BY SIMILARITY.  
 FT DISULFID 1926 1943 BY SIMILARITY.  
 FT DISULFID 1945 1954 BY SIMILARITY.  
 FT MOD\_RES 1591 1591 BY SIMILARITY.  
 FT MOD\_RES 1810 1810 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1036 1036 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1501 1501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1565 1565 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1741 1741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1827 1827 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1900 1900 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2024 2024 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2043 2043 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2061 2061 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2323 2323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2345 2345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 2923 AA; 317447 MW; 382757915158BD8 CRC64;  
 Query Match 100.0%; Score 15545; DB 1; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSPPATGVLPPLPPLLLLLPPLGPOVGCRLSGRSGSGACAPMGWLCPS 60  
 DB 1 MSPPAVGVLPPPPPLLLLLPPLGPOVGCRLSGRSGSGACAPMGWLCPS 60  
 QY 61 SASNLVLTYSRCRDAGTELTVGHVPHDGLRVWCPSBAHPLPPAPBCCPMSCRLLGIG 120  
 DB 61 SASNLVLTYSRCRDAGTELTVGHVPHDGLRVWCPSBAHPLPPAPBCCPMSCRLLGIG 120  
 QY 121 GHLSPPGKLTLPBEPHCLKAPRLRCQSCGLADAPGLRAEBSPEBELGRRRRNNVTAQ 180  
 DB 121 GHLSPPGKLTLPBEPHCLKAPRLRCQSCGLADAPGLRAEBSPEBELGRRRRNNVTAQ 180  
 QY 181 FQPPSYQATVPENQAGTVASLRALDPDEGEAGRLTYTMDLFDERSNQFSLDPVTGA 240  
 DB 181 FQPPSYQATVPENQAGTVASLRALDPDEGEAGRLTYTMDLFDERSNQFSLDPVTGA 240



241 VTTAELEBETKSTHVRVTAADHGMPPRSALATITLVTOTNDHPVPEOEYKESIRE 300  
241 VTTAELEBETKSTHVRVTAADHGMPPRSALATITLVTOTNDHPVPEOEYKESIRE 300  
301 NLEVEYEVLTVAATGDAPPNANILYRLLBGSQSPSEYFELDPSSGYIATRGPPDREB 360  
301 NLEVEYEVLTVAATGDAPPNANILYRLLBGSQSPSEYFELDPSSGYIATRGPPDREB 360  
361 ESYQULTEASDQGRDPGRSTTAAVFLSEVEDNDNAPQFSEKRYVVOYREBVTGAPYL 420  
361 ESYQULTEASDQGRDPGRSTTAAVFLSEVEDNDNAPQFSEKRYVVOYREBVTGAPYL 420  
421 VTAASRDGNSAVVHYYSIMSGNARQFYLDAGTGLDVVSPLDYETTKETLRVAOQGG 480  
421 VTAASRDGNSAVVHYYSIMSGNARQFYLDAGTGLDVVSPLDYETTKETLRVAOQGG 480  
481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESYPLGYLVHVAOIAADAGDARL 540  
481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESYPLGYLVHVAOIAADAGDARL 540  
541 EYRLAGVGHDPFTINNGTGMISVAELDRBEVDYFSGVEARDHGTALTAASVSVT 600  
541 EYRLAGVGHDPFTINNGTGMISVAELDRBEVDYFSGVEARDHGTALTAASVSVT 600  
601 LDVNDNNFTPOPEYTVRLNEDAAVGSVWYVSAIDRAHSHYITQITSGTNRKPSITS 660  
601 LDVNDNNFTPOPEYTVRLNEDAAVGSVWYVSAIDRAHSHYITQITSGTNRKPSITS 660  
661 QSGGGLVSIALPLDYKLEROYVLAVTASDGTRODPAQIVNVTADNTHRPVFOSSHYYN 720  
661 QSGGGLVSIALPLDYKLEROYVLAVTASDGTRODPAQIVNVTADNTHRPVFOSSHYYN 720  
721 VNEBRPAGTAVLISATBEDTGENARITYFMEDSIPQFRIDAGTAVTQAELEDEYVS 780  
721 VNEBRPAGTAVLISATBEDTGENARITYFMEDSIPQFRIDAGTAVTQAELEDEYVS 780  
781 YTLATADNNGIPKSDPTTYLEIIVNDVNDNAPOGLRBSYOGSYVEDPPTSVLQJAT 840  
781 YTLATADNNGIPKSDPTTYLEIIVNDVNDNAPOGLRBSYOGSYVEDPPTSVLQJAT 840  
841 DRDSGLNRRVFTYFQGGDGDGFVSESTSGIVRTLRLEBENVAQYVLAAYADKMP 900  
841 DRDSGLNRRVFTYFQGGDGDGFVSESTSGIVRTLRLEBENVAQYVLAAYADKMP 900  
901 ARTPEVTVTVLDVNDNPPVFEODEFVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
901 ARTPEVTVTVLDVNDNPPVFEODEFVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
961 VEGNIPVFOJLDPGELTALVDLYEDRPREVLYTQTSAPLVSRATVHRLDDRNDP 1020  
961 VEGNIPVFOJLDPGELTALVDLYEDRPREVLYTQTSAPLVSRATVHRLDDRNDP 1020  
1021 PVLGNFELLFNNVYVNRSSSPGAGIGRAYPAHDPDISLTYSPFERGKELSVILNASTG 1080  
1021 PVLGNFELLFNNVYVNRSSSPGAGIGRAYPAHDPDISLTYSPFERGKELSVILNASTG 1080  
1081 ELKISRALDNNRPLBAIVSVISDGVHSVTAQCALRTVITIDEMLTHSITLRLDMSER 1140  
1081 ELKISRALDNNRPLBAIVSVISDGVHSVTAQCALRTVITIDEMLTHSITLRLDMSER 1140  
1141 FLSPILGLFIOAVAAATLATPDHVVVVRVORDTAPRGHILNLSVQOPRGPGGPF 1200  
1141 FLSPILGLFIOAVAAATLATPDHVVVVRVORDTAPRGHILNLSVQOPRGPGGPF 1200  
1201 PSEBLOERLYLNRSLTALISAORVLPFDNCLREPCENYRCSVILFDSAPFIASS 1260  
1201 PSEBLOERLYLNRSLTALISAORVLPFDNCLREPCENYRCSVILFDSAPFIASS 1260  
1261 VLFPRPIHVGGJRCPCPGFTGDIYETEVDCYSPCGPHRCRSRGYTCICRDGYTG 1320  
1261 VLFPRPIHVGGJRCPCPGFTGDIYETEVDCYSPCGPHRCRSRGYTCICRDGYTG 1320  
1321 EHCEVSASRGCTPGVCNKGCTCVNLLVGFKCPCDPSGDFEKPVCQYVTTSPFAHSFTTF 1380

1321 EHCEVSASRGCTPGVCNKGCTCVNLLVGFKCPCDPSGDFEKPVCQYVTTSPFAHSFTTF 1380  
1381 RGLRORHFTLALSPATKEDGILLVNGFRBKEDPALAVIOBOVOLTFSAGSSTTVS 1440  
1381 RGLRORHFTLALSPATKEDGILLVNGFRBKEDPALAVIOBOVOLTFSAGSSTTVS 1440  
1441 PFVPGVSDQGMHTVOLKTYNKPALAGTGLPOGSEBQKVAVVYDGDGTVALRFGSYLG 1500  
1441 PFVPGVSDQGMHTVOLKTYNKPALAGTGLPOGSEBQKVAVVYDGDGTVALRFGSYLG 1500  
1501 NYSQAQGTQGSKSLDITGPLLGGVPLPESFPVRMQPVGCNRLQYDSHIMAD 1560  
1501 NYSQAQGTQGSKSLDITGPLLGGVPLPESFPVRMQPVGCNRLQYDSHIMAD 1560  
1561 PIANGTVPGCPAKCNCDSNTCHNGTCVNOQMAFSCBCLPFGGKSCAOEMANPOHFL 1620  
1561 PIANGTVPGCPAKCNCDSNTCHNGTCVNOQMAFSCBCLPFGGKSCAOEMANPOHFL 1620  
1621 GSSLVAMHGLSLPISQPMYLSLMFRTRQADGVLLQAITRGKSTITTLQREGHVLSEGT 1680  
1621 GSSLVAMHGLSLPISQPMYLSLMFRTRQADGVLLQAITRGKSTITTLQREGHVLSEGT 1680  
1681 GLQASSRLRPRGRANDGMHQAOLALGASGPGHAILSPYGOQBARGNLGPRLHGLHS 1740  
1681 GLQASSRLRPRGRANDGMHQAOLALGASGPGHAILSPYGOQBARGNLGPRLHGLHS 1740  
1741 NITVGGIPGAGVARGFRGCLQGVRSVDPTEGNSLDPBSHGSINVEOGCSLDPDCSN 1800  
1741 NITVGGIPGAGVARGFRGCLQGVRSVDPTEGNSLDPBSHGSINVEOGCSLDPDCSN 1800  
1801 PCPANSYCSNDWDSYSCSDPGYGDNCITVCDLNPCEHOSVCTRKPSAPHYTCBCPN 1860  
1801 PCPANSYCSNDWDSYSCSDPGYGDNCITVCDLNPCEHOSVCTRKPSAPHYTCBCPN 1860  
1861 YLGPYCESTRIDQCPKRWKMHPTCGPCNCVSKGFPDCKTSGECHKENHTRPSSPT 1920  
1861 YLGPYCESTRIDQCPKRWKMHPTCGPCNCVSKGFPDCKTSGECHKENHTRPSSPT 1920  
1921 CLLCDCTPGLSRVCPEDPGQCPCKPGVIGROCDRCNPFABVTNCEVYDSCRAI 1980  
1921 CLLCDCTPGLSRVCPEDPGQCPCKPGVIGROCDRCNPFABVTNCEVYDSCRAI 1980  
1981 BAGIMWPRTRFGJLPAAPCEKSGFATAVRHCDHRGMLPMLFNCTSIITFSBLKGAERL 2040  
1981 BAGIMWPRTRFGJLPAAPCEKSGFATAVRHCDHRGMLPMLFNCTSIITFSBLKGAERL 2040  
2041 QRNESGLDSRSQLALLLNATQHTAGYGSVKAAYQATRLALHESYORGLSATQ 2100  
2041 QRNESGLDSRSQLALLLNATQHTAGYGSVKAAYQATRLALHESYORGLSATQ 2100  
2101 DVHFTBELLVSGALDPTANKRMELIQQTEGGTAMLLQHYEYASALQNMRTYLSPF 2160  
2101 DVHFTBELLVSGALDPTANKRMELIQQTEGGTAMLLQHYEYASALQNMRTYLSPF 2160  
2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGQPPLETTVILPESVFRETPPVVRPAG 2220  
2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGQPPLETTVILPESVFRETPPVVRPAG 2220  
2221 PGEAQBEEELARORHREBELSOGBAVASVITYTLAGLLPHANDPDRSRIRVPRPINT 2280  
2221 PGEAQBEEELARORHREBELSOGBAVASVITYTLAGLLPHANDPDRSRIRVPRPINT 2280  
2281 PVSISVHDEBELLPALDQVTVQFRLLETERETPICVFNMHSILVSGTGMARGCE 2340  
2281 PVSISVHDEBELLPALDQVTVQFRLLETERETPICVFNMHSILVSGTGMARGCE 2340  
2341 VVRNBSHVSQCNHMTSPAVLMDVSRRENGEILPLKTYLTVVALGVTALALLTFPELTL 2400  
2341 VVRNBSHVSQCNHMTSPAVLMDVSRRENGEILPLKTYLTVVALGVTALALLTFPELTL 2400  
2401 LRLRSNQGIRBNLRAALGLAQVFLIGNOMDLPACTVIVILLHFLVLCFESNALLE 2460

Db 2401 LRLKRSNQHGRIRNRLTRALGLAQLVFLGLGINQADLPACTVIALILHFLYLCFSSMALLR 2460  
 Qy 2461 ALHLYALTEVRDVTNTPMRFYMLGWGVPATITGLAVGLDPGEGNPDCEMLSTVDTLLI 2520  
 Db 2461 ALHLYALTEVRDVTNTPMRFYMLGWGVPATITGLAVGLDPGEGNPDCEMLSTVDTLLI 2520  
 Qy 2521 WSPAGPVAFVNSVFLYIYAARASCAARQGFEEKGVPAGLOPSPAVILLISATWLLAL 2580  
 Db 2521 WSPAGPVAFVNSVFLYIYAARASCAARQGFEEKGVPAGLOPSPAVILLISATWLLAL 2580  
 Qy 2581 LSVNSDTLFLHYLPATNCIOGPFILSYVYLKSKYRKALKACSRKSPDPALTTKSTL 2640  
 Db 2581 LSVNSDTLFLHYLPATNCIOGPFILSYVYLKSKYRKALKACSRKSPDPALTTKSTL 2640  
 Qy 2641 TSSYNCPSPYADRLQPYGDSAGSLHSTSRSGKSPSYIPLLRBSALNPGQPGGLG 2700  
 Db 2641 TSSYNCPSPYADRLQPYGDSAGSLHSTSRSGKSPSYIPLLRBSALNPGQPGGLG 2700  
 Qy 2701 DPQSLFLREGDDQHDPTSDSLSLBDDQSGSYASTHSDSEEBEEREAAPFGEQG 2760  
 Db 2701 DPQSLFLREGDDQHDPTSDSLSLBDDQSGSYASTHSDSEEBEEREAAPFGEQG 2760  
 Qy 2761 WDSLPGARLPLHSTPKDGGPGKAPWPGDPTTAKSSGNGAPERRLRNGDALSR 2820  
 Db 2761 WDSLPGARLPLHSTPKDGGPGKAPWPGDPTTAKSSGNGAPERRLRNGDALSR 2820  
 Qy 2821 EGSGLPLPGSSAOPHKGILKCKCLPTISEKSLRLPLBQCTGSSRSGSSRSGGAPP 2880  
 Db 2821 EGSGLPLPGSSAOPHKGILKCKCLPTISEKSLRLPLBQCTGSSRSGSSRSGGAPP 2880  
 Qy 2881 RPPRRQSLQQLNGVWPIAMSIAGVYDESSGSEFLFFNFLLH 2923  
 Db 2881 RPPRRQSLQQLNGVWPIAMSIAGVYDESSGSEFLFFNFLLH 2923

RESULT 2  
 CLR2\_MOUSE STANDARD; PRT; 2920 AA.  
 ID CLR2\_MOUSE  
 AC Q9ROM0; Q99K6; Q9Z2R4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Caderin EGF LMG seven-pass G-type receptor 2 precursor (Flamingo 1) (Mfam11).  
 GN CELSR2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99418630; PubMed=10490098;  
 RA Uenai T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L., Takeichi M., Uemura T.;  
 RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Fz1212ed.";  
 RL Cell 98:585-595(1999).  
 RN (2)  
 RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.  
 RX MEDLINE=20253755; PubMed=10790539;  
 RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;  
 RT "Chromosomal localization of Celstr2 and Celstr3 in the mouse; Celstr3 is a candidate for the floppy (fip) lethal mutant on chromosome 9.";  
 RL Mamm. Genome 11:392-394(2000).  
 RN (3)  
 RP SEQUENCE OF 2014-2920 FROM N.A.  
 RC TISSUE-Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Bhat N.K., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F., Datchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carrincci P., Prange C., Raha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boase S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Meyers R.M., Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smailus D.E., RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN (4)  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=21839555; PubMed=11850187;  
 RA Tissot F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;  
 RT "Developmental expression profiles of Celstr (Flamingo) genes in the mouse.";  
 RL Mech. Dev. 112:157-160(2002).  
 CC -1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed in the CNS and in the developing CNS.  
 CC -1- DEVELOPMENTAL STAGE: Predominantly expressed in the developing CNS. The emerging dorsal root ganglia and cranial ganglia. In the CNS, expression is uniform along the rostrocaudal axis. During gastrulation, it is expressed within the anterior neural ectoderm. At E10, expression is strong in the ventricular zones (VZ) in all sectors of the brain, and lower in the marginal zones (MZ). Between E12 and E15, expression is prominent in the brain. It is strong in VZ, lower in MZ, except in telecephalic MZ where it is predominant. The intensity is higher in all VZ, and lower in differentiating fields than in VZ, except in the cerebral hemispheres, and to a lesser extent in the cecum and cerebellum. A weak expression is also observed in the fetal lungs, kidney and epithelia. In the newborn and postnatal stages, expression remains restricted to the VZ as well as in migrating and postmitogenic cells throughout the brain.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: Contains 9 cadherin domains.  
 CC -1- SIMILARITY: Contains 8 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin G-EGF-like domain.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AB028499; BAA84070.1; -  
 DR EMBL, AF031573; AAC68837.1; -  
 DR EMBL, BC005499; AAH05499.1; -  
 DR HSSP, P00740; 1EDM.  
 DR MGD, MGI:1858235; Celstr2.  
 DR GO, GO:0016021; C:Integral to membrane, ISS.  
 DR GO, GO:0004930; F:G-protein coupled receptor activity, ISS.  
 DR InterPro, IPR000152; Asx\_hydroxyl.  
 DR InterPro, IPR002126; Cadherin.  
 DR InterPro, IPR000742; EGF 2.  
 DR InterPro, IPR006209; EGF-like.  
 DR InterPro, IPR000832; GPCR\_secretin.  
 DR InterPro, IPR001879; hormn\_receptor.  
 DR InterPro, IPR002049; laminin\_EGF.  
 DR InterPro, IPR001791; laminin\_G.  
 DR InterPro, IPR000203; PKD\_cys\_rich.  
 DR Pfam, PF00002; 7cm\_2; 1.

[illegible]

Query Match	Similarity	Score	DB	Length	2920;
Beat Local	Similarity	94.4%;	Pred. No. 0;		
Matches 2760;	Conservative	54;	Mismatches 105;	Indels 5;	Gaps 2;
FT DISULFID	1356	1366	BY SIMILARITY.		
FT DISULFID	1579	1590	BY SIMILARITY.		
FT DISULFID	1584	1599	BY SIMILARITY.		
FT DISULFID	1601	1610	BY SIMILARITY.		
FT DISULFID	1792	1803	BY SIMILARITY.		
FT DISULFID	1798	1818	BY SIMILARITY.		
FT DISULFID	1820	1829	BY SIMILARITY.		
FT DISULFID	1833	1844	BY SIMILARITY.		
FT DISULFID	1838	1856	BY SIMILARITY.		
FT DISULFID	1858	1867	BY SIMILARITY.		
FT DISULFID	1868	1900	BY SIMILARITY.		
FT DISULFID	1890	1907	BY SIMILARITY.		
FT DISULFID	1909	1922	BY SIMILARITY.		
FT DISULFID	1925	1937	BY SIMILARITY.		
FT DISULFID	1927	1944	BY SIMILARITY.		
FT DISULFID	1946	1955	BY SIMILARITY.		
FT MOD RES	1592	1592	BY SIMILARITY.		
FT CARBOHYD	486	486	HYDROXYLATION (POTENTIAL).		
			N-LINKED (GLCNAC. .) (POTENTIAL).		
Query Match	94.4%;	Score 14657.5;	DB 1;	Length 2920;	
Beat Local	Similarity	94.4%;	Pred. No. 0;		
Matches 2760;	Conservative	54;	Mismatches 105;	Indels 5;	Gaps 2;
QY 1	MRSATGVPPLPTPEPPLLLLLLLLLLPPPLIGDVGPCRSIGSRGRSSGACAPMGWCP	60			
DB 1	MRTAAAPLPPTPLPLLLLLLLLLLPPSPLLGDVGPCRSIGSGRSSGACAPMGWCPA	60			
QY 61	SASUMLYTCRCADAGTELGHVPHHDLGRWCPCSEBAHPLPPAPEGCPWSCRLIGIG	120			
DB 61	SASUMLYTCRCSESGIELGHVPHHDLGRWCPCSEBAHPLPPSEGGCPWSCRLIGIG	120			
QY 121	GHLSPOGKTLPEBHPCLKAPRLRCOSCKLAQAPGLRAGERSPEBSIGRRKKNVTAPO	180			
DB 121	GHLSPOGKTLPEBHPCLKAPRLRCOSCKLAQAPGLRAGERSPEBSIGRRKKNVTAPO	180			
QY 181	POPSPYATVPENOPACTPVASLRALDPDGBGGRLETTMDALPDSRNPFSLDPVTGA	240			
DB 181	POPSPYATVPENOPACTPVASLRALDPDGBGGRLETTMDALPDSRNPFSLDPVTGV	240			
QY 241	VTTAEELDRKTSKTHVGRVTAQDHGMRRSALATLTLVTDTHDPVFEEQSEKSLRE	300			
DB 241	VTTAEELDRKTSKTHVGRVTAQDHGMRRSALATLTLVTDTHDPVFEEQSEKSLRE	300			
QY 301	NLEVGIVLTVRATDGDAPPNANILYRLLEGSGSGSEVFEIDPRSGVITRGPVDRREV	360			
DB 301	NLEVGIVLTVRATDGDAPPNANILYRLLEGAGDSPEDAFEIDPRSGVITRGPVDRREV	360			
QY 361	ESYOLATFEASDQDRPDPRSTTAAVFLSVDDNDMNPQSEKKYVYQVREDVPGAPVL	420			
DB 361	ESYOLATFEASDQDRPDPRSTTAAVFLSVDDNDMNPQSEKKYVYQVREDVPGAPVL	420			
QY 421	VTAASDRDKGSNAVHYHSIMSGNARGFYLDQAQALDVAVSLDYETTKETKYLTVRAQDG	480			
DB 421	VTAASDRDKGSNAVHYHSIMSGNARGFYLDQAQALDVAVSLDYETTKETKYLTVRAQDG	480			
QY 481	RPLSNVSGIVYVYQVINDIND-NAPIVSSTPFOATVLESVPGLTVLHYQALIDADGDNAR	539			
DB 481	RPLSNVSGIVYVYQVINDIND-NAPIVSSTPFOATVLESVPGLTVLHYQALIDADGDNAR	540			
QY 540	LEYLAVGVADPFPTINNGTGMISVAALDREBEVDYFSFVEARNDGTPLATASASVAT	599			
DB 541	LEYLAVGVADPFPTINNGTGMISVAALDREBEVDYFSFVEARNDGTPLATASASVAT	600			
QY 600	VLDVNDNPPFTPOPEYTVRLNEBAVAGTSVYVYSAVDRDAHSYITQTISGTRNRPST	659			
DB 601	VLDVNDNPPFTPOPEYTVRLNEBAVAGTSVYVYSAVDRDAHSYITQTISGTRNRPST	660			
QY 660	SOSGGGLVSTALPLDYLKEROYLAATASGTQDPAQOIVYVNTDANTHRPVSQSHYV	719			
DB 661	SOSGGGLVSTALPLDYLKEROYLAATASGTQDPAQOIVYVNTDANTHRPVSQSHYV	720			
QY 720	NVNEEDPAGTIVVLIATYEDTGSEANRIYFMEDSIPOFRIDADTGAVTTQAEIDYEDV	779			
DB 720	NVNEEDPAGTIVVLIATYEDTGSEANRIYFMEDSIPOFRIDADTGAVTTQAEIDYEDV	779			

Db 721 NGNEDRPAQTVVLIISATDEDTGENARITTFMEDSIPQFRIDGDTGAVTTQAEIJDYQV 780  
Qy 780 SYTLAIIPARDNGIPOKSDITTYLEILVNDVMDNAPQFLRDSYGOSVYEDVPPTSYLQISA 839  
Db 781 SYTLAIIPARDNGIPOKSDITTYLEILVNDVMDNAPQFLRDSYGOSVYEDVPPTSYLQISA 840  
Qy 840 TDRDSGLNGRVPFTYPOGDDGDDGFIVESTSIGVRLRLRDLRENVAYQVLAAYDKGMP 899  
Db 841 TDRDSGLNGRVPFTYPOGDDGDDGFIVESTSIGVRLRLRDLRENVAYQVLAAYDKGMP 900  
Qy 900 PARTPEVTVTVLDVNDNPPVFEDEDFVVEENSPGLAVAYATADPDEGTNAQIMYO 959  
Db 901 PARTPEVTVTVLDVNDNPPVFEDEDFVVEENSPGLAVAYATADPDEGTNAQIMYO 960  
Qy 960 IVEGNIPEVFOQDIFSGELTALVDLDYEDRPEYVLVQATSAPLVSRATVHRLDNDN 1019  
Db 961 IVEGNIPEVFOQDIFSGELTALVDLDYEDRPEYVLVQATSAPLVSRATVHRLDNDN 1020  
Qy 1020 PVLVGNFELLFNNVYTNRSSSPFGAIGRVPAHDPDISDLTYSPERGNELSLVLNAST 1079  
Db 1021 PVLVGNFELLFNNVYTNRSSSPFGAIGRVPAHDPDISDLTYSPERGNELSLVLNAST 1080  
Qy 1080 GELKLSALNNPPLKALMSVLSDGYHSTAQCALAVTTITDBMLTHSTITLLEDMSPR 1139  
Db 1081 GELKLSALNNPPLKALMSVLSDGYHSTAQCALAVTTITDBMLTHSTITLLEDMSPR 1140  
Qy 1140 RFLSPILGLFIOAVATLATPPDHVVVFNQORDACGHIILNVSLSVQPPGPGGPP 1199  
Db 1141 RFLSPILGLFIOAVATLATPPDHVVVFNQORDACGHIILNVSLSVQPPGPGGPP 1200  
Qy 1200 LPSEBDLOERLYLNRSLITLTAISAQVLPDPDNICLREPCENYMCVSVLAFDSSAPFIASS 1259  
Db 1201 LPSEBDLOERLYLNRSLITLTAISAQVLPDPDNICLREPCENYMCVSVLAFDSSAPFIASS 1260  
Qy 1260 SVLFRPHVPGJRCRCRCPGPTGDTYDCEYDLCYSRPGHGRCSRSEGGYTCCLGDCYT 1319  
Db 1261 SVLFRPHVPGJRCRCRCPGPTGDTYDCEYDLCYSRPGHGRCSRSEGGYTCCLGDCYT 1320  
Qy 1320 GEHCENVSABSGRCCTPGVCKNGGTCVNLVAGAFKDCPSGDPREKPCVATTTRSPAFHSFT 1379  
Db 1321 GEHCENVSABSGRCCTPGVCKNGGTCVNLVAGAFKDCPSGDPREKPCVATTTRSPAFHSFT 1380  
Qy 1380 FRGIRQRFHTTALSPATKRDGILLYNGRPNRKHDPVALLEVIOEVOULTFSAGESTTY 1439  
Db 1381 FRGIRQRFHTTALSPATKRDGILLYNGRPNRKHDPVALLEVIOEVOULTFSAGESTTY 1440  
Qy 1440 SPFVPGSVSDGOMHTVOLKTYNKPILGQGTGLPOGSEBQKXAVVTVBSCDGVALLRGSYV 1499  
Db 1441 SPFVPGSVSDGOMHTVOLKTYNKPILGQGTGLPOGSEBQKXAVVTVBSCDGVALLRGSYV 1500  
Qy 1500 GNTSCAAQCTQGSGKSLDLTGPIILGVPDLBPESFVRMRQFVGCWRLQVDSRHIDMA 1559  
Db 1501 GNTSCAAQCTQGSGKSLDLTGPIILGVPDLBPESFVRMRQFVGCWRLQVDSRHIDMA 1560  
Qy 1560 DFLANCTVPGCPAKKXVCSNCTCANGCTCYNOMDAPSCCEPLJFGGKSQAQEMANPOHF 1619  
Db 1561 DFLANCTVPGCPAKKXVCSNCTCANGCTCYNOMDAPSCCEPLJFGGKSQAQEMANPOHF 1620  
Qy 1620 LGSLSLVAMHGLSLPIGOWTILMFTTRQADGVTLQAITRGRSTITTLQLEBGMVLSVEG 1679  
Db 1621 LGSLSLVAMHGLSLPIGOWTILMFTTRQADGVTLQAITRGRSTITTLQLEBGMVLSVEG 1680  
Qy 1680 TGLQASSLRLPERGANDGDWHHAQALGASGPGHAILSPDYQOQARBNLAGRLHGLHL 1739  
Db 1681 TGLQASSLRLPERGANDGDWHHAQALGASGPGHAILSPDYQOQARBNLAGRLHGLHL 1740  
Qy 1740 SNTTGGIIPBPAAGVANGFRGCLQGVAVSDTPBGSVNLDSHGBSINVEGCSLLPDPDS 1799  
Db 1741 SNTTGGIIPBPAAGVANGFRGCLQGVAVSDTPBGSVNLDSHGBSINVEGCSLLPDPDS 1800  
Qy 1800 NPCEPANSYCNMDMDSYSCSDPGYGYGDNCTNVCDLNCEHQSVCTRKXPAAPHGYTCBPP 1859  
Db 1801 NPCEPANSYCNMDMDSYSCSDPGYGYGDNCTNVCDLNCEHQSVCTRKXPAAPHGYTCBPP 1860

Qy 1860 NYLGPYCETRIDOPCPRGMWGHPTGCPNCNDVSKGFPDPCNKTSGECHKENHYRPPGSP 1919  
Db 1861 NYLGPYCETRIDOPCPRGMWGHPTGCPNCNDVSKGFPDPCNKTSGECHKENHYRPPGSP 1920  
Qy 1920 TGLCDVPTGSLSRVCDDEDOCPCKPVGIGROCDRCNPPAVYTNCGEAVNYDSCPA 1979  
Db 1921 TGLCDVPTGSLSRVCDDEDOCPCKPVGIGROCDRCNPPAVYTNCGEAVNYDSCPA 1980  
Qy 1980 IEAGIWPPTRGALPAAPCPGSGFVAVHCDERHGMPLPULFNCTSTFSELGFAAR 2039  
Db 1981 IEAGIWPPTRGALPAAPCPGSGFVAVHCDERHGMPLPULFNCTSTFSELGFAAR 2040  
Qy 2040 LQNRSGLDGSGRQOLALLRRAOTHTAGYFGSDVKVAVQOLATRLLAHSTORGLSAT 2099  
Db 2041 LQNRSGLDGSGRQOLALLRRAOTHTAGYFGSDVKVAVQOLATRLLAHSTORGLSAT 2100  
Qy 2100 QVHFTENILRVGSALLDPAANRHHWELIQTEGTRAMLLQHYEAYASALAOHNRHTYLS 2159  
Db 2101 QVHFTENILRVGSALLDPAANRHHWELIQTEGTRAMLLQHYEAYASALAOHNRHTYLS 2160  
Qy 2160 FTIVTPNIVISVRLDKGNFAGKLPYRYALRGEOPDLETTVIIPESVFRETPPVVRA 2219  
Db 2161 FTIVTPNIVISVRLDKGNFAGKLPYRYALRGEOPDLETTVIIPESVFRETPPVVRA 2220  
Qy 2220 GPGEAQEPPELARBORRHPBELSQGEAVASVIIYRTLAGLPHNYDPDKRSLRVPKRPITN 2279  
Db 2221 GPGEAQEPPELARBORRHPBELSQGEAVASVIIYRTLAGLPHNYDPDKRSLRVPKRPITN 2280  
Qy 2280 TVVVISVHDEBELPRALDKVTVYQFRLLEBETKCPVCMWNSILVSGTGMASRG 2339  
Db 2281 TVVVISVHDEBELPRALDKVTVYQFRLLEBETKCPVCMWNSILVSGTGMASRG 2340  
Qy 2340 EYVFRNESHVSQCNHMTSFAVLMDSVRRENGBILPLKTLVYVAGVTLAALLTFPFLT 2399  
Db 2341 EYVFRNESHVSQCNHMTSFAVLMDSVRRENGBILPLKTLVYVAGVTLAALLTFPFLT 2400  
Qy 2400 LRLIRSNQHGIRNMTALAGLAQVFLGIGNOQDLPRACVIAIILHFLYCTFSMALL 2459  
Db 2401 LRLIRSNQHGIRNMTALAGLAQVFLGIGNOQDLPRACVIAIILHFLYCTFSMALL 2460  
Qy 2460 EALHLYRALTETVDVNTGPMRYTMYLGVNPAFITIGLAVGLDPRGYGNDPCWLSIYDTL 2519  
Db 2461 EALHLYRALTETVDVNTGPMRYTMYLGVNPAFITIGLAVGLDPRGYGNDPCWLSIYDTL 2520  
Qy 2520 IWSFAGPVAFVMSVFLYITLAAASCAAQOGFEKSGVSGIOPFAVTLILSATWILIA 2579  
Db 2521 IWSFAGPVAFVMSVFLYITLAAASCAAQOGFEKSGVSGIOPFAVTLILSATWILIA 2580  
Qy 2580 LLSVNSDITLLFHYLPATNCIOGPPIFLSYVVLSEKVRKALKLACSRRKSPDPALTTKST 2639  
Db 2581 LLSVNSDITLLFHYLPATNCIOGPPIFLSYVVLSEKVRKALKLACSRRKSPDPALTTKST 2640  
Qy 2640 LTRSYNCPBPYADGELYOPYGBSAGLSHSTSGSOSBYIPFLIRBSLTPNGVPPDL 2699  
Db 2641 LTRSYNCPBPYADGELYOPYGBSAGLSHSTSGSOSBYIPFLIRBSLTPNGVPPDL 2700  
Qy 2700 GDPGLFLTEGQOQDPTDSDSLLEDDOGSASTHSSDSEEBEEREEAFAFGEO 2759  
Db 2701 GDPGLFLTEGQOQDPTDSDSLLEDDOGSASTHSSDSEEBEEREEAFAFGEO 2760  
Qy 2760 GWDLSLGPAGERLPLHSTPKOGGPGPKAPMGDFGTAKESSGNGAPBEERLRENGDLS 2819  
Db 2757 GWDLSLGPAGERLPLHSTPKOGGPGPKAPMGDFGTAKESSGNGAPBEERLRENGDLS 2816  
Qy 2820 REGSLGPIPGSGAOGHKLKKKCLPTLSKXSLRLPLBOCTGSRGSSASBEGSGGPP 2879  
Db 2817 REGSLGPIPGSGAOGHKLKKKCLPTLSKXSLRLPLBOCTGSRGSSASBEGSGGPP 2876  
Qy 2880 PRPPPOSLOEQNLGVMPDIAMSIRKAGTVDBDSGSGEPFLFNFHL 2923  
Db 2877 PRPPPOSLOEQNLGVMPDIAMSIRKAGTVDBDSGSGEPFLFNFHL 2920



Query Match 71.4%; Score 11097; DB 1; Length 2144;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 2062; Conservative 35; Mismatches 47; Indels 4; Gaps 1;  
SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;  
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 966 966 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1249 1249 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1286 1286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1548 1548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1570 1570 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;  
Query Match 71.4%; Score 11097; DB 1; Length 2144;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 2062; Conservative 35; Mismatches 47; Indels 4; Gaps 1;  
QY 776 EDVSYTLATARDNGIPKSDTTYLEILVNDVNDNAPOFLRDSYQGSVEYEDVPFTSVL 835  
DB 1 EDVSYTLATARDNGIPKSDTTYLEILVNDVNDNAPOFLRDSYQGSVEYEDVPFTSVL 60  
QY 836 QISATRDSGLNGRVFTYFQGGDGDGFVSTSGIVRTLRRLRENAQYVLRAYVD 895  
DB QISATRDSGLNGRVFTYFQGGDGDGFVSTSGIVRTLRRLRENAQYVLRAYVD 120  
QY 896 KGNPAPATPMEVYTVTVLVNDNPPVFEODEPVDVENSPIGLAVAVATTDDEGNAQ 955  
DB KGNPAPATPMEVYTVTVLVNDNPPVFEODEPVDVENSPIGLAVAVATTDDEGNAQ 180  
QY 956 IMVQIVENGIPEVFOLDIPFSGELTALVDLDYEDRPEVLYVQATSAPLVRATVAVELD 1015  
DB IMVQIVENGIPEVFOLDIPFSGELTALVDLDYEDRPEVLYVQATSAPLVRATVAVELD 240  
QY 1016 RNDNPPVLANPELIFNNYVTRNSSSPFGAIGRVPAPDPI DSBLYSPFERGELSTVLL 1075  
DB RNDNPPVLANPELIFNNYVTRNSSSPFGAIGRVPAPDPI DSBLYSPFERGELSTVLL 300  
QY 1076 NASTGELKLSRALDNNPLRAIMSVLSDGVHSTACALRVITIDEMTHSITLLED 1135  
DB NASTGELKLSRALDNNPLRAIMSVLSDGVHSTACALRVITIDEMTHSITLLED 360  
QY 1136 MSBERPLSLPLGLFIQAVATLATPDPHVVFVNOQRTDAPGGHILNVSLVQCPPGG 1195  
DB MSBERPLSLPLGLFIQAVATLATPDPHVVFVNOQRTDAPGGHILNVSLVQCPPGG 420  
QY 1196 GPPPLPEBDLQERLYLNRSLITLTAISQORVLPFDNITCLAREPCENYMCVSLRPSGAP 1255  
DB GPPPLPEBDLQERLYLNRSLITLTAISQORVLPFDNITCLAREPCENYMCVSLRPSGAP 480  
QY 1256 IASSSVFRLPHVGLRCRCPGFTGDCETEDVLCYSRPCGPHGCRSREGYTLCR 1315  
DB IASSSVFRLPHVGLRCRCPGFTGDCETEDVLCYSRPCGPHGCRSREGYTLCR 540  
QY 1316 DGYTGBCEVASARSGRCTPGVCKNGTCVNLVGVGFCDCPSGFEKPCQVTRSPAR 1375  
DB DGYTGBCEVASARSGRCTPGVCKNGTCVNLVGVGFCDCPSGFEKPCQVTRSPAR 600  
QY 1376 SFTIFRGLRQRFHTTALSPATKRDGLLYNGFRNEKDFVALVETQEOVOLTFSAGES 1435  
DB SFTIFRGLRQRFHTTALSPATKRDGLLYNGFRNEKDFVALVETQEOVOLTFSAGES 660  
QY 1436 TTVTSFVPGVSGQVHTVQKTKYKPLLGQNGLPQSPSOKAVAVTVGCDTVALRF 1495  
DB TTVTSFVPGVSGQVHTVQKTKYKPLLGQNGLPQSPSOKAVAVTVGCDTVALRF 720  
QY 1496 GSVLGNYSCAAQGTQGSKSLDLTGPILLGGVLDLPSPPVNRQFVGCNRLQVDSRH 1555  
DB GSVLGNYSCAAQGTQGSKSLDLTGPILLGGVLDLPSPPVNRQFVGCNRLQVDSRH 780  
QY 1556 IDMAFDIANNCTVGCRAKKNVCDNNTCHNGCTCVNQMADAPSGCEPLGPGKSCAQEMAN 1615

DB 781 VDMAFDIANNCTVGCRAKKNVCDNNTCHNGCTCVNQMADAPSGCEPLGPGKSCAQEMAN 840  
QY 1616 PQHFLASSLVAMHGLSLPISQPMYLSMERTQADGVLLQAITRGSTITTLQREHVL 1675  
DB PQHFLASSLVAMHGLSLPISQPMYLSMERTQADGVLLQAITRGSTITTLQREHVL 900  
QY 1676 SVEGTGLQASSLRLEGRANDDMHQAOLASGGPGHAIISFDVQGOARAGNLGRLLH 1735  
DB SVEGTGLQASSLRLEGRANDDMHQAOLASGGPGHAIISFDVQGOARAGNLGRLLH 960  
QY 1736 GLHLSNTVGIIPGAPGAVARFGCLQGVRSVDPTEGVNSLDPSSHESINVEQCSLPD 1795  
DB GLHLSNTVGIIPGAPGAVARFGCLQGVRSVDPTEGVNSLDPSSHESINVEQCSLPD 1020  
QY 1796 PCDSNPCPNASYCSNDMDSYSCSDPGYVGDCTVNCIDLPCEHOSACTRKSAPRGYTC 1855  
DB PCDSNPCPNASYCSNDMDSYSCSDPGYVGDCTVNCIDLPCEHOSACTRKSAPRGYTC 1080  
QY 1856 ECEPNVLPYCESTRIDQPCFRGMWGHPTGPNCDVSKGFPDPCNKTSGECHKENHYP 1915  
DB ECEPNVLPYCESTRIDQPCFRGMWGHPTGPNCDVSKGFPDPCNKTSGECHKENHYP 1140  
QY 1916 PSSPTCLCDYCPYGLSLRVCDPBDGQCEKVGVIQRQCDRCNDPFAVYTTNGCEVYDS 1975  
DB PSSPTCLCDYCPYGLSLRVCDPBDGQCEKVGVIQRQCDRCNDPFAVYTTNGCEVYDS 1200  
QY 1976 CPRAIENAGIWMPTRGFLPAAAPCPKSGRGTAVRCHDERGMLPPLFNCSTISSEKNG 2035  
DB CPRAIENAGIWMPTRGFLPAAAPCPKSGRGTAVRCHDERGMLPPLFNCSTISSEKNG 1260  
QY 2036 FAERLQRNESGLDSRQOLALLRNATQHTAGVSGDVKVAVOLATRLIAESTORGFG 2095  
DB FAERLQRNESGLDSRQOLALLRNATQHTAGVSGDVKVAVOLATRLIAESTORGFG 1320  
QY 2096 LSATQDVHTENULRAYGALLDPANKRMELIQTBGTAUMLQHYEAYASLAQNMRT 2155  
DB LSATQDVHTENULRAYGALLDPANKRMELIQTBGTAUMLQHYEAYASLAQNMRT 1380  
QY 2156 YLSPFTVFNVISVVRLDKGNPAGAKLPRYBALRGEPDLETTVLIPESVPRETPV 2215  
DB YLSPFTVFNVISVVRLDKGNPAGAKLPRYBALRGEPDLETTVLIPESVPRETPV 1440  
QY 2216 VBPAGGEAQBEBELARRQRRPBLSGEAVASVITRTLAGLPHNYDPDRSLRVPKR 2275  
DB VBPAGGEAQBEBELARRQRRPBLSGEAVASVITRTLAGLPHNYDPDRSLRVPKR 1500  
QY 2276 PIIINTPVVISVHDBELLPRALDYPVTVQFRLBTEERTKEICVFMWNSILVSGTGWMS 2335  
DB PIIINTPVVISVHDBELLPRALDYPVTVQFRLBTEERTKEICVFMWNSILVSGTGWMS 1560  
QY 2336 ARGCEVFNRESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLTVAGLTLAALLTF 2395  
DB ARGCEVFNRESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLTVAGLTLAALLTF 1620  
QY 2396 FFLTLARILRSNOHIGIRNLTALGLAOLVFLGINQADLPACTVIALTLHFLYCTES 2455  
DB FFLTLARILRSNOHIGIRNLTALGLAOLVFLGINQADLPACTVIALTLHFLYCTES 1680  
QY 2456 WALLEMLHLYALTRVDVNTGPNRYTLMGVGVAFTIGLAVGDDPEYGNPDPCMLSI 2515  
DB WALLEMLHLYALTRVDVNTGPNRYTLMGVGVAFTIGLAVGDDPEYGNPDPCMLSI 1740  
QY 2516 YDTLIMSFPAPVAFVAMSVFVYIILASBASCAAQGFKEKPKVSGLSRSTVLLLSAT 2575  
DB YDTLIMSFPAPVAFVAMSVFVYIILASBASCAAQGFKEKPKVSGLSRSTVLLLSAT 1800  
QY 2576 WLLIALISVNSDTLLFHTYLFATONCTIQPPIFLSYVVLSEVRKALKACSRSPDPALT 2635  
DB WLLIALISVNSDTLLFHTYLFATONCTIQPPIFLSYVVLSEVRKALKACSRSPDPALT 1860  
QY 2636 TKSITLTSVNCSPVADGRLVQPYGDSAGSLSTSRSGSOSPXYPLLRSSALNPGOG 2695



```

Db 1861 TKTSLTSSNCPSPYADGRUYOYXGDSAGSLHNSHSGSQSPYIPFLRREBTNPGOV 1920
Qy 2696 PPELGDPGSLFLEGQDQDHPDSDSDLEDDQSGSVASTHSDSEEEEEEAAAF 2755
Db 1921 PPELGDPGSLFMEGQAQOHDPDSDSDLEDDQSGSVASTHSDS9---EEEEEAAAF 1976
Qy 2756 PPEQGDSDLLGPAEAEPLPLHSTKDGPGGKAPWQDPTTAKESGNGAPERLRENG 2815
Db 1977 PEGQGDSDLLGPAEAEPLPLHSTKDGPGGKAPWQDPTTAKESGNGAPERLRENG 2036
Qy 2816 DALSRGSLGPIPLGSSGAOPHKGLKKKCLPTISEKSLILRLPLRQCTGSSRGSSASRGR 2875
Db 2037 DALTRGSLGPIPLGSPETQPHKGLKKKCLPTISEKSLILRLPLEQGTGSSRGSSASRGR 2096
Qy 2876 GGGPPPPPPROSLQEQQLNGVPIAMSIKACTVDEDSGSEFLPFNFILH 2923
Db 2097 NGPPPPPPPPROSLQEQQLNGVPIAMSIKACTVDEDSGSEFLPFNFILH 2144

RESULT 4
CLRI MOUSE STANDARD; PRT; 3034 AA.
AC 035161;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF IAG seven-pass G-type receptor 1 precursor.
GN CELSR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN NM (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9077689; PubMed=9858697;
RA Hadjantonakis A.-K., Formstone C.J., Little P.F.R.;
RT "Celser1 is an evolutionarily conserved seven-pass transmembrane
RT receptor and is expressed during mouse embryonic development.";
RL Mech. Dev. 78:91-95(1998).
(2)
RP TISSUE SPECIFICITY.
RX STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=97480720; PubMed=9339365;
RA Hadjantonakis A.-K., Sheward W.J., Hartmar A.J., de Galan L.,
RA Hoovers J.M.N., Little P.F.R.;
RT "Celser1, a neural-specific gene encoding an unusual seven-pass
RT transmembrane receptor, maps to mouse chromosome 15 and human
RT chromosome 22qter.";
RL Genomics 45:97-104(1997).
(3)
RP DEVELOPMENTAL STAGE.
RX MEDLINE=21839555; PubMed=11850187;
RA Tisler P., De-Backer O., Goffinet A.M., Lambert de Rouvroir C.A.;
RT "Developmental expression profiles of Celser (Flamingo) genes in the
RT mouse.";
RL Mech. Dev. 112:1157-160(2002).
-1- FUNCTION: Receptor that may have an important role in cell/cell
-1- SIGNALING during nervous system formation.
-1- SUBCELLULAR LOCATION: Integral membrane protein.
-1- TISSUE SPECIFICITY: Expressed in the brain, where it is localized
-1- principally in the ependymal cell layer, choroid plexus and the
-1- area postrema. Also found in spinal chord and in the eye.
-1- DEVELOPMENTAL STAGE: First detected at E6. Predominantly expressed
-1- in the developing CNS, the emerging dorsal root ganglia and
-1- cranial ganglia. In the CNS, expression is uniform along the
-1- rostrocaudal axis. During gastrulation, it is expressed in the
-1- vicinity of the primitive streak, and becomes predominant in that
-1- area at late gastrulation. At E10, detected in ventricular zones
-1- (VZ), but not in marginal zones (MZ), and weakly in other
-1- structures. Between E12 and E15, a high expression is present in
-1- the VZ in all brain areas. No expression in differentiated
-1- neuronal fields. In the newborn and postnatal stages, expression
-1- remains restricted to the VZ. Also found weakly in fetal lungs,

```

```

CC kidney and epithelia.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF031572; AAC68836.1; -.
DR PIR, T14119; T14119.
DR MGD, MGI:1106883; Celser1.
DR HSSP, P00749; IURK.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; horum_receptor.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR001791; laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00053; laminin_EGF; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00303; GPS_1.
DR SMART; SM00303; GPS_1.
DR SMART; SM00008; Horns; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS0010; ASX_HYDROXYL; 2.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 9.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50221; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE; PS50227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE; PS50025; Lam_G_DOMAIN; 2.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 29
FT CHAIN 30 3034
FT DOMAIN 21 2484
FT TRANSMM 2485 2505
FT TRANSMM 2506 2516
FT TRANSMM 2517 2537
FT DOMAIN 2538 2542
FT TRANSMM 2543 2563
FT TRANSMM 2564 2587
FT TRANSMM 2588 2608
FT TRANSMM 2609 2625
FT TRANSMM 2626 2646
FT TRANSMM 2647 2670
FT TRANSMM 2671 2691

```

PT DOMAIN 2692 2694 EXTRACELLULAR (POTENTIAL).  
 PT TRANSMEM 2695 2715 7 (POTENTIAL).  
 PT DOMAIN 2716 3034 CYTOPLASMIC (POTENTIAL).  
 PT DOMAIN 261 368 CADHERIN 1.  
 PT DOMAIN 369 474 CADHERIN 2.  
 PT DOMAIN 475 580 CADHERIN 3.  
 PT DOMAIN 581 702 CADHERIN 4.  
 PT DOMAIN 703 804 CADHERIN 5.  
 PT DOMAIN 805 907 CADHERIN 6.  
 PT DOMAIN 908 1014 CADHERIN 7.  
 PT DOMAIN 1015 1116 CADHERIN 8.  
 PT DOMAIN 1121 1239 CADHERIN 9.  
 PT DOMAIN 1318 1376 EGF-LIKE 1, CALCIUM-BINDING.  
 PT DOMAIN 1378 1414 EGF-LIKE 2, CALCIUM-BINDING.  
 PT DOMAIN 1418 1456 EGF-LIKE 3, CALCIUM-BINDING.  
 PT DOMAIN 1457 1661 LAMININ G-LIKE 1.  
 PT DOMAIN 1664 1700 EGF-LIKE 4, CALCIUM-BINDING.  
 PT DOMAIN 1704 1885 LAMININ G-LIKE 2.  
 PT DOMAIN 1887 1922 EGF-LIKE 5, CALCIUM-BINDING.  
 PT DOMAIN 1923 1961 EGF-LIKE 6, CALCIUM-BINDING.  
 PT DOMAIN 1962 1994 EGF-LIKE 7, CALCIUM-BINDING.  
 PT DOMAIN 1996 2031 EGF-LIKE 8, CALCIUM-BINDING.  
 PT DOMAIN 2037 2070 LAMININ EGF-LIKE.  
 PT DOMAIN 2070 2475 GPS.  
 PT DOMAIN 2475 2678 POLY-LRU.  
 PT DISULFID 1322 1333 BY SIMILARITY.  
 PT DISULFID 1327 1364 BY SIMILARITY.  
 PT DISULFID 1366 1375 BY SIMILARITY.  
 PT DISULFID 1382 1393 BY SIMILARITY.  
 PT DISULFID 1387 1402 BY SIMILARITY.  
 PT DISULFID 1404 1413 BY SIMILARITY.  
 PT DISULFID 1422 1433 BY SIMILARITY.  
 PT DISULFID 1427 1443 BY SIMILARITY.  
 PT DISULFID 1445 1455 BY SIMILARITY.  
 PT DISULFID 1668 1679 BY SIMILARITY.  
 PT DISULFID 1673 1688 BY SIMILARITY.  
 PT DISULFID 1690 1699 BY SIMILARITY.  
 PT DISULFID 1891 1902 BY SIMILARITY.  
 PT DISULFID 1896 1911 BY SIMILARITY.  
 PT DISULFID 1913 1922 BY SIMILARITY.  
 PT DISULFID 1926 1937 BY SIMILARITY.  
 PT DISULFID 1931 1949 BY SIMILARITY.  
 PT DISULFID 1951 1960 BY SIMILARITY.  
 PT DISULFID 1968 1981 BY SIMILARITY.  
 PT DISULFID 1983 1993 BY SIMILARITY.  
 PT DISULFID 2000 2015 BY SIMILARITY.  
 PT DISULFID 2002 2018 BY SIMILARITY.  
 PT DISULFID 2020 2030 BY SIMILARITY.  
 PT MOD\_RES 1681 1681 HYDROXYLATION (POTENTIAL).  
 PT MOD\_RES 1904 1904 HYDROXYLATION (POTENTIAL).  
 PT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 793 793 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1129 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1228 1228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1264 1264 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1591 1591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1655 1655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 1994 1994 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2118 2118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2137 2137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2144 2144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2155 2155 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2160 2160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2452 2452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PT CARBOHYD 2538 2538 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3034 AA; 330477 MW; EFF38180AF5ED8A8 CRC64;  
 Query Match 57.7%; Score 8974.5; DB 1; Length 3034;  
 Best Local Similarity 57.3%; Fred. No. 0;  
 Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;  
 1 MRSPTAGVLPPTPPPLLLLLLLLPPPLIGQVGCRCISGRGSGS-----S 48  
 124 LRSNAGALRSP-----AVRSVPLGDL--CFRAGGGAASLSVLEAITNPPA 172  
 49 GACAPM-----GWC--PSSASINMLYTSRCRDAGTELGHVPHDGLRWCPSSSA 99  
 173 CSCPPVAGTGCRRGPCLRPGSABRLVCAIGRAAGA-----VWV----- 213  
 100 HPLPAPGCGWCSCLLDIGHLSPQGLTLPEHPCAKAPRLRCQCKLAQAPLBRAG 159  
 214 -----ELVIQATSGTPSESPV-SPSL---LNLSP--RAG 243  
 160 --ERSPEBSLGGRRKNVTAPQPPSYQATVPENQAPGTVAASIRAI DPDEGEAGRL 217  
 244 VVRS-----RGGSGTSPPQPLPSYQVGVPEHNPAGTAVIELRAHPDEGDAGRLS 296  
 218 YTMALPDRSNQFSLDPVTGAVTAEELDRBTSYVFRVYADHGMRRSALATLT 277  
 297 YQWEALPDRSNGYFLIDATGATVARSILDRBTOTHLKYSADVHGSPPRSAAVYTV 356  
 278 LVTDPNDHPVPRQOEKESLRNLEVGREVTVAATGDAPNANILYRLGSGSGPS 337  
 357 TVSDTNDSHPVEQSYRERIRENLEVGREVTVAATGDAPNANILYRLGSGSGPS 413  
 338 EYFELIDPRSGVLRTRGPVDRREEVSYQLTVEASDQGRDPGRSTTAAVELSVEDNDNAP 397  
 414 -VEIDARSGVTRAVVDREBAEYQLLVEANDQGRNGLSASATVHIVVEDENDNP 472  
 398 QSEKRYTVQVQEDVTGAPVLRVTSADRKSNVHVSINSGNARQGYLDAQCALD 457  
 473 QSEKRYTVQVQEDVAVNTAVRVQATDRDQGNIAHVSIVSGNKGFPYHLSJSGSLD 532  
 458 VVSPFLDYETTKETLRVRAQDGRPPLSNVSGLVTVQVLDINDNAP1FSTPFOATVLS 517  
 533 VINPLDFEALIRYTLIRAKQDGRPPLINSGLVSVQVLDVNDNAP1FVSSFPQAVLNP 592  
 518 VPLGYLVLFVQALIDADAGNARLELYLAGVH-----DPEFTINNGTG 560  
 593 VPLGSHVLIHQAVDADAGNARLYRLVDTASTIVGSSVDSBNPASADPFOJHNSG 652  
 561 WTSVAEELREEDVPSRGVEARDHGTPLTASASVSVTVLVNNNPFPTQPEYTVRLN 620  
 653 WTVCAELDRBEVEHYSFGEZAVDHGSPMSSASASVTVLVNDNDPMFTQPVYELRLN 712  
 621 EDAAVGTSVVTVSAVDRDAHVSYYTQITSGNTRNRFSTISQSGGGLVSLALPLDYLERQ 680  
 713 EDAAVSSVTLIRARDRDANSVITQITGNTNRNFPALSSQGGGLITALLPLDYLERQ 772  
 681 YVLAVTASDGTRODTRAQIVVNTVDANTHRPVQSHYTVVNVNEDRPAGTTVVLISATDE 740  
 773 YVLAVTASDGTSHRQVIVNTDANTHRPVQSHYTVVSEDEPVGFSITATISATDE 832  
 741 TGENAITYPMEDSLPQRFIDADTGAVTQALDYEDQVSYLATARPNGIPQSDTTY 800  
 833 TGENAITYVLDPVQFIDPTGTITMTLEDVEDQAAYLATIQAQNGIPQSDTIS 892  
 801 LELIVNDVNDNAPQFLRDSYQGVVEDVPFTSVLQISATDRDGLNGRVFTYFGGDDG 860  
 893 LELILIDANDNAPRFLRDPYQGVVEDAPPSVSVQVASTDRDGSNGRLVLTPOGGDDG 952  
 861 DGDFTYESTSGVTRTLRDLRENVAYVLRAYAVDKMP--PARTMEVTVTVLVNDNRP 919  
 953 DGFYFYEPTSGVIRQORLDRENVAVYVLMALAVRGSPNPASAGIVSVTVLDINDNP 1012  
 920 VFEQDFDVFEVENSIGLAVARVATDPDEGNQAIQVIEGNIPEVFOIDIFSGEJT 979  
 1013 VFEKDLLELVEENSPVGSVVARIRANDPBGPMQIITQIVEGNVFVFOJDLISGLLR 1072

QY 980 ALVLDYEDPERYLVIQATSAPIVSHATVHLLDNDNPVYAGNEELLFNNYVNRSS 1039  
 DB 1073 ALVELPEVARDVWLVOATSAPIVSHATVHLLDNDNPVYAGNEELLFNNYVNRSS 1132  
 QY 1040 SPFGGAIIRVAPADPDSDSLTYSFERGNEELSLVLNASTGELKLSALDNNRPLRLANS 1099  
 DB 1133 SPFGGAIIRVAPADPDSDSLTYSFERGNEELSLVLNASTGELKLSALDNNRPLRLANS 1192  
 QY 1100 VLVSVDGSHVTAOCALRVITITDEMLTHSITLRLDMSPERPLSLGLFIOAVATLAT 1159  
 DB 1193 VVSVDGSHVTAOCALRVITITDEMLTHSITLRLDMSPERPLSLGLFIOAVATLAT 1252  
 QY 1160 PPHVAVVFNQRTDAPGHIILNVSLSVGPDPGGPPLPSEDIQERLYNRSLLTAL 1219  
 DB 1253 TKXDIFFVFNQNDTV--SSNINLVTFESALPGGTG--RFPSEDIQERLYNRSLLTAL 1309  
 QY 1220 SAORVLPFDNICTLRPCENYMRVSVLRFDDSAFPLASSVLFRPHVPGILRCPCPG 1279  
 DB 1310 SAORVLPFDNICTLRPCENYMRVSVLRFDDSAFPLASSVLFRPHVPGILRCPCPG 1369  
 QY 1280 FTGDYCEBTEVDLCYSRCPHGRCSRREGYTCLCDGYTGEHCDEVASRGRCTPGVCKN 1339  
 DB 1370 FTGDYCEBTEVDLCYSRCPHGRCSRREGYTCLCDGYTGEHCDEVASRGRCTPGVCKN 1429  
 QY 1340 GGTICVNLVGGFCDCPSGDPEKPYCQVTRSPHASFITRGLRORPHFTLALSPATKE 1399  
 DB 1430 GGTICVNLVGGFCDCPSGDPEKPYCQVTRSPHASFITRGLRORPHFTLALSPATKE 1489  
 QY 1400 RDLGLLYNGRFRKNDPVALEVIOBOVQLTFESAGESTTVSPPVPGVSGQCHHTYQLK 1459  
 DB 1490 RDLGLLYNGRFRKNDPVALEVIOBOVQLTFESAGESTTVSPPVPGVSGQCHHTYQLK 1549  
 QY 1460 YNPKLLGQTCGLPOGSEBQKAVVTVVDCDGTVALRFEVSLGANSCAAQGTGSSKSLDL 1519  
 DB 1550 YNPKLLGQTCGLPOGSEBQKAVVTVVDCDGTVALRFEVSLGANSCAAQGTGSSKSLDL 1609  
 QY 1520 TGPLLLGGVNDLDESPVVRKROVYVGCERNLQVDSRHIDMADPIANNGTVPCCAKNVC 1579  
 DB 1610 TGPLLLGGVNDLDESPVVRKROVYVGCERNLQVDSRHIDMADPIANNGTVPCCAKNVC 1669  
 QY 1580 SNTCHNGGTVCNOMDAPSCPLGFGKSCAOMANPOHFLGSSLVAMHGLSLPTIQPMY 1639  
 DB 1670 SNTCHNGGTVCNOMDAPSCPLGFGKSCAOMANPOHFLGSSLVAMHGLSLPTIQPMY 1729  
 QY 1640 LSLMFTROADGVLLQAITRGSTITLQLEBGMVLSVBEGTGLQASSLRLPERANDGM 1699  
 DB 1730 LSLMFTROADGVLLQAITRGSTITLQLEBGMVLSVBEGTGLQASSLRLPERANDGM 1789  
 QY 1700 HHAQLALASGGPCH-----AIIISPDYGOORABGNLGPRLHGLHLSNITVVGIPGAGV 1754  
 DB 1790 HHAQLALASGGPCH-----AIIISPDYGOORABGNLGPRLHGLHLSNITVVGIPGAGV 1848  
 QY 1755 ARGFRGLCGVRVSDTPEGVNSLDPHGESINTEOCGLPDPDSCNPRANSTCSNDMS 1814  
 DB 1849 ARGFRGLCGVRVSDTPEGVNSLDPHGESINTEOCGLPDPDSCNPRANSTCSNDMS 1908  
 QY 1815 YSCSCDPGYGDNCTVNCDLNPECOSVCTRKPSAPHGTYCECPVYLAGYCESTRIDOPC 1874  
 DB 1909 YSCSCDPGYGDNCTVNCDLNPECOSVCTRKPSAPHGTYCECPVYLAGYCESTRIDOPC 1968  
 QY 1875 PRGMWGHPTGPNCDVSKGFPDCKNTSGECHKENHTRPPGSPICLLCDYCYTGSLSR 1934  
 DB 1969 PRGMWGHPTGPNCDVSKGFPDCKNTSGECHKENHTRPPGSPICLLCDYCYTGSLSR 2028  
 QY 1935 VCPBEGQCCPKRGVIGROCDRCNDPFAVTTNGCEVNVDSGRATEAGIMPRTRFGLP 1994  
 DB 2029 VCPBEGQCCPKRGVIGROCDRCNDPFAVTTNGCEVNVDSGRATEAGIMPRTRFGLP 2088  
 QY 1995 AAAPCPKSGFATVRECHDERGMLPMLFNCTSTTSEBLKGFALRLQRNESGLDSGRSOQ 2054  
 DB 2089 AAAPCPKSGFATVRECHDERGMLPMLFNCTSTTSEBLKGFALRLQRNESGLDSGRSOQ 2148

QY 2055 LALLANATOHTAGYGSVDKVAAYOLATRELLAHSTORGSLSANODVHTENTLAVGSA 2114  
 DB 2149 LALLANATOHTAGYGSVDKVAAYOLATRELLAHSTORGSLSANODVHTENTLAVGSA 2208  
 QY 2115 LLDTANKRWEILIOOTEGTATLLOHYEAYASALAKONMHTLSPPTIYTPNIVISVRL 2174  
 DB 2209 LLDTANKRWEILIOOTEGTATLLOHYEAYASALAKONMHTLSPPTIYTPNIVISVRL 2268  
 QY 2175 DKGNFAGAKLPRYEAALRGOPDLETTVILPSSVF-----RBTIPPYVR-----PAGE 2221  
 DB 2269 DKGNFAGAKLPRYEAALRGOPDLETTVILPSSVF-----RBTIPPYVR-----PAGE 2328  
 QY 2222 GEAQEBEELIARORRHPHLSQGBAASVYIYTTAGLLPHNYDPDKRSIYVRRPIINTP 2281  
 DB 2329 GEAQEBEELIARORRHPHLSQGBAASVYIYTTAGLLPHNYDPDKRSIYVRRPIINTP 2388  
 QY 2282 VYSISYHDBELLPRALDKPVVOORLLTERTRPICYFMNHSILVSGTGSARGCYV 2341  
 DB 2389 VYSISYHDBELLPRALDKPVVOORLLTERTRPICYFMNHSILVSGTGSARGCYV 2448  
 QY 2342 VERNESHVSCQCNHMTSPAVLMDVSRRENGILPLKTLTYVALGYLALLLTFPPLTL 2401  
 DB 2449 VERNESHVSCQCNHMTSPAVLMDVSRRENGILPLKTLTYVALGYLALLLTFPPLTL 2508  
 QY 2402 RILRSNQGIRNLTALGLAQVFLGQADLPACTVIALHLHPLYLCTFSMALTEA 2461  
 DB 2509 RILRSNQGIRNLTALGLAQVFLGQADLPACTVIALHLHPLYLCTFSMALTEA 2568  
 QY 2462 LHLRALTRVROVNTGPMRFFYMLGWPAPITGLANGIDPEGYGNPDCMSTYDTLIW 2521  
 DB 2569 LHLRALTRVROVNTGPMRFFYMLGWPAPITGLANGIDPEGYGNPDCMSTYDTLIW 2628  
 QY 2522 SPAGVAFVSMVSVLYLAARASCAORQGEKKGPVSGLOPSFAVILLLSATVILAL 2581  
 DB 2629 SPAGVAFVSMVSVLYLAARASCAORQGEKKGPVSGLOPSFAVILLLSATVILAL 2688  
 QY 2582 SVNSDTLLFHYLPATCNCIQGFIFLSYVLSKEYRAK-LACSRKPSDPALTTKST- 2639  
 DB 2689 SVNSDTLLFHYLPATCNCIQGFIFLSYVLSKEYRAK-LACSRKPSDPALTTKST- 2748  
 QY 2640 LITSYNCSPPYADG--RLYQPGDSAGLSHRSRSGKSPSVIIPFLRESMLNQCQGP 2697  
 DB 2749 LITSYNCSPPYADG--RLYQPGDSAGLSHRSRSGKSPSVIIPFLRESMLNQCQGP 2798  
 QY 2698 GLADPGSLFT--EGQPOHDPDTSDDSLSLBDQGSYASLTHSDSEBEESEEAAP 2756  
 DB 2799 GLADPGSLFT--EGQPOHDPDTSDDSLSLBDQGSYASLTHSDSEBEESEEAAP 2853  
 QY 2757 GEQWDSLLGPAERLPLHSTPK-DGPGPGKAPMPGD--FGTTAKE----- 2800  
 DB 2854 GEQWDSLLGPAERLPLHSTPK-DGPGPGKAPMPGD--FGTTAKE----- 2903  
 QY 2801 -----SSGNAPBERLRENDALSRBGLSLPLGSSAOP--HKGLIKKCL--PTIS 2848  
 DB 2904 -----SSGNAPBERLRENDALSRBGLSLPLGSSAOP--HKGLIKKCL--PTIS 2958  
 QY 2849 EK--SLLRLPLFOCTGSSRGGS--SASESGRGPP-----RPPPOSIOLEQLGVMP 2898  
 DB 2959 EK--SLLRLPLFOCTGSSRGGS--SASESGRGPP-----RPPPOSIOLEQLGVMP 3015  
 QY 2899 AMSIAGVTDEDSGSE 2915  
 DB 3016 AMSIAGVTDEDSGSE 3032

RESULT 5  
 CLRL HUMAN STANDARD; PRT, 3014 AA.  
 AC 09NY06; 09Y722; 09BM05; 09Y506; 09Y526;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Caderlin Bgf IAG seven-pass G-type receptor 1 precursor (Flamingo



[illegible]

Query Match	56.3%	Score 8754	DB 1	Length 3014
Best Local Similarity	55.8%	Pred. No. 0		
Matches 1703	Conservative 420	Mismatches 687	Indels 240	Gaps 39

  

QY	12	TEPPPLLLLLLLLLLPPPLLDQVGCRLSGRGGSSGACAPMGWLCFSSASNMLVYTSR	71
Db	57	TPAPRRELL-----DVGDGDLAIG-RVSGAGRPV-----PLQVRLV-----	93
QY	72	CRDAGTELYGVPHHDGLRVWCPSBAHPLPAPBEGCPMSCRLLGGHLSPOGKLT	131
Db	94	ARSAFTALSRLL-----RAKTHLP-----GCGARARLCGTGARLC-GALCF	133
QY	132	P-----EEHPCLKA-----PLRFGS-----CKLAQAPG-	155
Db	134	FVPGGCAAAQSHSALAAPTLPACRCPPRPBRCPPRCPICTLPGGGSVRLRLCALARRAGA	193
QY	156	LRAG-----ERSPESLGGRRKRNVTAP-----QFOPSYQAT	189
Db	194	VRVGLAEAAATAGTPSSASPSPPPLPPLPLFEARAGPARARAGTSGRGSIAFPMNYQVA	253
QY	190	VPEHQPACTPVASLRALIDPDGEAGRLKLYTMDALFDSNSNOFSLDPYTGAVTTAAEELDR	249
Db	254	LEFNEPAGTLLQLHAAYTIGBEEBSVSYMEGLDEBSRGVFRIDSTAGAVSTDSVDR	313

QY	250	EFKSHYRVAJAOHGMREBSALMTLTLVMDTDNDHDFVQOQEKESLRETLFVGEVTL	309
Db	314	ETKETHVLRVAVDSTDPKSAFTYTLVLVDXNDHDSVFPEQSEFRERREKREMLEVGEVL	373
QY	310	TVRAITDGDAPNPANILYRLLEBGSQGSPEVEIDIPRSGVITREGRFVDEEYESYOLTVEA	369
Db	374	TIRASDRDSPINAMLRXYVLGGAM----DVFOINSSGVSVSTRALVDREBAEYOLIVEA	429
QY	370	SDQGRDPRBRTTAAVPLFVSVEDDNDNAPQSEBEKRYVQVREDVTPGAAYLRYTASDRQK	429
Db	430	NDQGNPNPPLSATYATVYILEVEDENDNPFQSEBOYVWQVPEDVGMLTVLRQAIDRQO	489
QY	430	SNAAVYHSGNAGROFYLDAGTGALDVSPFLYETKETELTLRVAODGGAPPLSNVSG	489
Db	490	QMAIHHSIILSGNVAAGPYLHSLSGIIDLVPDLFEDVQKSLSKADQGRPRLINSNG	549
QY	490	LYTVQVLDINDNADIPVSTPFOATVLSBVGYLVLHVQAIDADGNARLEYRLA----	545
Db	550	VVSVOVLDVNDNEBIFVSSPFOATVLEVPPLGYEVVHQAVDADSGENARLHYELVDYAS	609
QY	546	----GVGH-----DPEPTINGGNTSVAAELDRBEVDFSPGVEARDHGPALTA	592
Db	610	TYLGGSGAGPNPAPTPDFPQJINSSGMITVCAELDRBEEHYHSPGYEADHDSPPMSS	669
QY	593	SASVSVTLVDVNDNPTFTQEPYTRLNEDBAVGSVYVYASAVDRDAHSVTTYQTSQNT	652
Db	670	STVSITVLVDNNDPVFTQPYTELRLNEDBAVQSSVLTLLQARDANASVTTYQLTQANT	729
QY	653	RNRSTISQSGGGVSLALPLDYLEEQYVLAATASGTRQDTQOIVYNTDANTHPRVF	712
Db	730	RNRALTSQSGGGGITTLALPLDYQEQOYVLAATASGTSHTAHVLAINTVANTHPRVF	789
QY	713	QSHYTVANVEDBRAGTTVVLISATBEDDGENAITYFMEDSIPOFRIDATGCAVTTQAE	772
Db	790	QSHYTVSVSEDRVGVGISATLSANDBDTGENAITYVIODPVQFRIDPDQSGMYTME	849
QY	773	LDVEDQVSYTLATARPNNGIPQKSDTTYLSILVNDVNDAPQPLRDSYQGSVEDPPEFT	832
Db	850	LDYENQVAYTLLTMAQONGIPQKSDTTYLTBLILDADNDAPOFLMDPYQSGIFEDAPPST	909
QY	833	SVLQISATDSDSGLANGVFYTPQGGDGGDGPFIYESGAYRTLRLRIDREVAQVYLRAY	892
Db	910	SILQVKSATDSDSGNGMLTTPQGGDGGDGPFIETPSGAYKTRKRLDREVAAYNLMAL	969
QY	893	AVDKAMP-PARTMVEVTVLVNDVNDPVEQEBDFVFEENSBDIGLAVARTATDDEG	951
Db	970	AVRGSFTPLSASVEIQTVITLIDINDNAPMEKELELFEENNPNVGSVAKIRANDDEG	1029
QY	952	TMAQIWAYIEGNIPVFOGLDIFSGELTALVDLDYERPRPYVYIOKTSAPLVBARAVHV	1011
Db	1030	PNAQIMQIAYIEGDMRHFQDLNLNGDIRANVELDFEYRREYVLAVOQTSAPLVSARAVHI	1089
QY	1012	RLLDNRNPNVPLGNFELFNNVYVYNNSSSPFGGAGIYVPAHPDPIISGLTYSFERGNEIS	1071
Db	1090	LVYQNDNPNVPLPDFOQLFNNVYVYNNKSNSPFTVIGICIPAHDPDVSGLSNMTYVQNGEIR	1149
QY	1072	LVYLNASTGELKLSRALDNNRPLEALINSVYVSGVNSVYTAQCALRTVITIDEMLTHTSITL	1131
Db	1150	LLLDPATYGELOQSRDLDDNNRPLEALMEVSVSGIHSVTAFCRLRYVITIDMDLTMSITV	1209
QY	1132	RLMEDSBERELSPGLGFIQAVAAATLATPPDHVVFVVOGDTAPAGHILNVSLSVQCP	1191
Db	1210	RLNEMSGEKLSPLMLAFVGVAAVSLTDDQDVFVRVQNDTBY-SSNIINVTYSALLPG	1268
QY	1192	GPQGGPPFLPSEDLQERLYNRSLLTAISAQRYLPPDNTCLBPCENYRQSVYLRFDS	1251
Db	1269	GVNRG--QFPSEBDOBOIYINRKLTTITSTORYLPPDNTCLBPCENYMKCVSVLRFDS	1326
QY	1252	SAPFIASSSVLPRIHVGGJLRCKCPGFGDVCETEVLD-CYASNPQGPBARCSREGGYT	1311
Db	1327	SAPPLSSTYVLPFRIHINGLRCKCPGFGDVCETETIDLCYDPCCANGCMSRBSGYT	1386
QY	1312	CLCRDGYTGEHCVASRGRCTPVCNKGNGTCVNLVGGFKCDQPSGDFEKPXYQVYTRS	1371

```

Db 1387 CEEFEDPTEHCEVDARSRCANGVCKNGTGVNLLIGFHCPCPGREYPERCEYTTTS 1446
Qy 1372 PFAHSFETRGRLRQRFHTLALSPATERDGLLYNGRFNEKADPALAEVIOQVOLTSS 1431
Db 1447 PFGQSVTRGRLRQRFHTLALSPATERDGLLYNGRFNEKADPALAEVIOQVOLTSS 1506
Qy 1432 AGESTTVBPVPGVSDQGMATVOLKYYNKPRLGGTGLPOGSEQKVAVTVVDCDTGV 1491
Db 1507 AGESTTVBPVPGVSDQGMATVOLKYYNKPRLGGTGLPOGSEQKVAVTVVDCDTGV 1566
Qy 1492 ALRFGSVLNGVSCAAGTGGSKSLDLNCPRLLLGGVPLPESFPYRMOQPGCMRNLQY 1551
Db 1567 AARFGDIGNVSCAAGTGGSKSLDLNCPRLLLGGVPLPESFPYRMOQPGCMRNLQY 1626
Qy 1552 DSRHIMADFIANNGTVPGCCPAKKNVCSNTCANGTCVNMOMAFSCEPGLRGSKCAQ 1611
Db 1627 DGNVDMAGFIANNGTVPGCCPAKKNVCSNTCANGTCVNMOMAFSCEPGLRGSKCAQ 1686
Qy 1612 EMANPQHLGSSLVANHGLSLPISQPMYLSIMFRTQADGVLLQAITRGRSTTLQJRBG 1671
Db 1687 AMPHPOLFSGESVVSNDLNIISVPMYLGIMFRTKEDSVLMEATSGGPTSFRLQILNN 1746
Qy 1672 HWMLSVEGVGLQASSLRBERGRANDGDMMHQAOL---GASGGRGHAL-LSPDYGGORAE 1727
Db 1747 YLQFVESHGSDVESVWLSGLRVTDBRMHLLLELNKXEDSEMKHLVTMTLDYGMDOVK 1806
Qy 1728 GNLGPRHLGHLHSNITVGGIPGPAAGVARGFGLQVRSVDPREGVNSLDPBSGSIYV 1787
Db 1807 ADIGMLPGLTVASVVGASBKVVRGFRGCMGVNMGITPVATLNNMNAKAVY 1866
Qy 1788 BQGSCLPDPGDSNCPANSTYCSNDMDSYSCSDPGYGDNCVTNCDLNPCEHOSVCTRXP 1847
Db 1867 KQGCDDVDDCTSSPCPNRSKCHAMEDYSCVCKGVTGINCVADACHLNCENMGALCVRSF 1926
Qy 1848 SAPHYTCEPRMYLGPYCESTRIDOPCRGMKHPFCGPNCDVSGSPBDCKTSGECH 1907
Db 1927 GSPQGVCEGPRHGYCENKLDLPCPRGMWGNPVCGRPHCAVSGPBDCKTSGECH 1986
Qy 1908 CKNHYPGSPGPTCLLDCYPTGSLRVCDPBDGQPCPKRGVIGROCDRCDNPAEVTN 1967
Db 1987 CKENYKLAQDTCPCDPCPHGSHRTCDMAQGCCKRGVIGROCDNCDNPAEVTN 2046
Qy 1968 GCEVNTDSCPRALFAGIWMPTRTFGRLPAAPCPKSGFRTAVNCHDEHGMPLPNTCTG 2027
Db 2047 GCEVNTDSCPRALFAGIWMPTRTFGRLPAAPCPKSGFRTAVNCHDEHGMPLPNTCTG 2106
Qy 2028 ITSEELKGAERLQHNESGLDSRQQLALLLNATQHTAGVSGDVKVAYOLATLLAH 2087
Db 2107 ISFVDLRANMEKLSRNBSTOVGABALQVRLRSATQHTGLTFLGNDVRTAYQLGHVQLH 2166
Qy 2088 ESTORGFGLSATODVHFTENLNVGSALLDTANKRMHLLIQTEGGTAMLLQHYEAYASA 2147
Db 2167 ESNQGGFDLAATODADPHEDVHNGSALLAPATRAMEOQRESEGTQALRLRLEBYFEN 2226
Qy 2148 LAQNMHTTISPTIYTPNIVISVRLDKNPNAGALPRYALRGSPDPLETTVILPSS 2207
Db 2227 VANNVRRTYLRPVITYANMLLVDFDKNFPGARPRDTHIEFPELBSVSFPRAD 2286
Qy 2208 VFR-----ETPPVVRPAG-----PGEAQBEBELARORRHPBELSOGSAVSVIYRT 2254
Db 2287 FFRPPEKSGPLRPRARRTTPTTRPGPETERBAPISRRRRPDDGCPAVALLVIYRT 2346
Qy 2255 LAGLLPHNTYDPRKSLRVPKRPINTPVVSVIHDEBELPRALDPRVTVQFLLTEER 2314
Db 2347 LQGLLPRYDPRRSLRLPRPIINTPVSTLYSEBAPLPRLEBPVLYEFALLVEER 2406
Qy 2315 TKRICYPMNHSILVSGGWSAGCEVFRNESHVSQCNHMTSPAVLMDVSRRENGEL 2374
Db 2407 TKRVCVFMNHSILVSGGWSAGCELSRNRHTVAQCSTHTASFAVLMDSRRENGEL 2466
Qy 2375 PLKTLFVALGVTLAALLTFPFLTLRLRSNOHGIRMLTAAAGLAQVFLGIGNAD 2434

```

```

Db 2467 PLKITYAASVLSLALLVAVFLSLVMLRSNHSIKHLAVALFLSQLVFVIGINQTE 2526
Qy 2435 LPPACTVAILLHFLYLCFSPMALLLEALHLYRALTEVDVNTGPMRFYMLGMGPAPIT 2494
Db 2527 NPFCLTVAILLHFLYLMSTPMTLVESLAVRMLTEVNIDTPKRFYVOMGIPALVT 2586
Qy 2495 GLAVGLDEBYGNPDFCWLSTYDILVNSPAGPVAFAVMSVFLYTLAARASCAORQFE 2554
Db 2587 GLAVGLDEBYGNPDFCWLSTYDILVNSPAGPVAFAVMSVFLYTLAARASCAORQFE 2646
Qy 2555 KKGPSVGLPSPAVYLLLSATMTALLSVNSDTHLPHLPAFCNIOGPFIPLSVYLSK 2614
Db 2647 KKGIVSLRTAFVLLLSATMTALLSVNSDTHLPHLPAFCNIOGPFIPLSVYLSK 2706
Qy 2615 EVKALK-LACSRKSPDPALTTKST-LTSSYNCPSPADG--RLYQPYGDSAGSLHSTS 2670
Db 2707 EVKRLKVLGGRKLHLEDSATRTATLTLRLSNCMTTGDDGPDMLRTLGSTASLDSIV 2766
Qy 2671 RS-----GKSQPSYIPFLRBSALNPGQPPGLDPSLPLEGDDQ 2713
Db 2767 RDEGIQKLGVSSGLVRGSHGPRDASIMPRSCD-----PFG----- 2802
Qy 2714 HDPDTSDDLSDBDQSGVASTHSSDSEBEEBEEBAAFPQSGWDSLPGAEKLP 2773
Db 2803 --HDSDSSELSL--DEQSSSTASHSSDSEDDGVAERK-----WDPARG-----A 2845
Qy 2774 LHSTPK-DGCGPQPKAPMPDFTGTTAKESGNGAP-----EERLRNCPAL 2818
Db 2846 VASTPKGAVANHPVAGMPDQSLAESDESDPSGRFLAVETKVSYLEHREGSGSHRGYP 2905
Qy 2819 SREBSLGLPSSAQP---HGILKKCL---PTISKS--SLRLPLEOCTGSSRGS 2869
Db 2906 PDBEGGAARLASQPPQKRGILKNKVTYPPPLTLTQTLKGRLEKLAQCEGPTSSR 2965
Qy 2870 ASEBSRGGP---PPRPPPSLOQLNCPWPLMSIRAGTVDBDSGSE 2915
Db 2966 TSSLSGGPDDCAITYKSPRBRGRDLNGV---ANAVRTGSHQADGSDSE 3012

RESULT 6
CLR3 HUMAN
ID CLR3 HUMAN STANDARD; PRT; 3312 AA.
AC 09NTQ7; 075092;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF lag seven-pass G-type receptor 3 precursor (Flamingo homolog 1) (hFml1) (Multiple epidermal growth factor-like domains 2) (Epidermal growth factor-like 1).
GN CELSR3 OR CDHF11 OR FM11 OR EGFL1 OR MEGP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

```





194 GSRRRVTAACGELMTGSKGQ-----BRATTSGAKTAPRNNCLFGASG----- 240  
QY EAHIFLPAPBEGCWMSCRLIGIGHLSPOGKLTLPESHPCIKAPRLKOS---CK-LAQA 153  
Db 241 -----SGPBLDSAPRTAKTAPASGS-APRESRTAPEAP-----KXMSRGLFCRFLPQR 230  
QY 154 PG-----LRAGSRPEESLGGRRKNVNTAPQPPSYQATVPENOPAGTPVASLRAT 206  
Db PGPRPPLPARPEARKTASNRARAFRRANRHPQPPQNYQTLVPEBNAAGTAVLRVAQ 350  
QY DPBGEAGLEXTMDALPDSRSNPFSLDPYTGAVNTAEELDPETKSTHFRTPADQHM 266  
Db DPDAEGEGRVTSIALAMNSRSLFSLIDPQSGLRTAALDRSMERHARTYAQDHS 410  
QY PRSALATLTLLVDTDHPVFEQOEKESLRENLEVEVLTAVRTDGDAPNNAILY 326  
Db PRSATMTAVTADRDHSPVEQOAYRETLRENVEGYPIQLRATDDBAPNNALRY 470  
QY 327 RL-EGSGSPSEVFEIDPRSGVITRGPVDBEEVSQUTVPSADQGRDPPRSTTAA 384  
Db RFGVPPARAAAAAFAFEIDPRSGLISTSGVDRHEMESYELVVEASDQGEPPRSATVR 530  
QY 385 VFLSVEDNNMNAPOPSKRYVVOVREBDVTGAPVLTATSDRKGSAVHVHSGSNAR 444  
Db VHTVLDENNAPQPSKRYVAQVREBDVPHVLTATDRDANGLVHYNISGNSR 590  
QY 445 GQFYLDAQTLADVSPBLDYETTKETTLARVADQGRPLSNVSGLTVOVLINDNAP 504  
Db 591 GHRAIDSLTGEIVQVADLPREAREVAKIRIADAGRPPLSNNTGLASIQVVDINHPI 650  
QY 505 FVSTPQATVLESVPLGLVLAHQVADADAGNARLEYRLAGVHDPEPTINNCTMISV 564  
Db 651 FVSTPQVSLXENAPLCHSVHIQVADADAGENARLEYSLGVAPDTPPINSATGAVSV 710  
QY 565 AABLDBREVDYFSGVGRARHGPALTAASVSSTVVDVNDNPTFQPEPTYRLMEDA 624  
Db 711 SGPLDRSVHETFEFVARHDSPPLSASAVTVDVNDNPEFTKEXYHRLMEDAA 770  
QY 625 VGTSVTVASAVDRDASVITYQITSGNTRNRFSTISQGGGLVSLALPLDYKLERQVLA 684  
Db 771 VGTSVSVTAVDRDANSAISYQITGNTNRFALSTQGGVGLVTLPLDYKERYKLV 830  
QY 685 VTSADGRTQTAQIVAVNTANTNRPVFOGSHYTVANNEDRPACTTVYLISATDEBTGEN 744  
Db 831 LTASDRALHCHCYHINITANTNRPVFOGSAHVSVMEDRPMSSTIVVLSASDDVGEN 890  
QY 745 ARTYFMEEDSIPOFRIDAGVTTQOABLDYEDVSTLAIATARDNGIPOKSPITYLEIL 804  
Db 891 ARTYFLEEDNLPQFRIDAGSGLTLOAPLDYEDVSTLAIATARDNGIPOKADITYVEVM 950  
QY 805 UNVDNNAPOFLRDSYOGSVYEDVPFTSVLOISATRDSGLNGRVFTEQGDGDGDF 864  
Db 951 UNVDNNAPOFVASHYTGIVSEBAPFTSVLOISATRDAMANGRVYTTQNEBDGDF 1010  
QY 865 IVBSTSGIVTLRLDREBNAQVYLAAYVDKMPPARPMETVTVTLVDNDNPPVEQD 924  
Db 1011 TIEPTSGIVTRRLDREBNAVSVELTAYAVDRGVPPLRTPTPSIQWVQDVNDNAPVPPAR 1070  
QY 925 BFDVFEVENSPIGLAVARVATDPBEGTNAQWQIYEGNIPIYFOLDITSGELTALVDL 984  
Db 1071 BFEFRVENSIVGSVQAITAVDPBEGNAHIMQIYEGNIPBLFQMDITSGELTALIDL 1130  
QY 985 DYEDRPYVLTQATSAPLVSRAVTVHRLDRNDNPPVLNFBLLFNNTYNSSSRPPG 1044  
Db 1131 DYBARQYVIVVQATSAPLVSRAVTVHRLVDQNDNSFVLNFFQLFNNYSNSDTPPSG 1190  
QY 1045 AIGRVAPRHPDISLSTYSPERGNELSLVLAASGTGKLSRALDNNRPBALMSVUSD 1104  
Db 1191 IIGIPIYDVPDVSHTLPSFPERGNELQLLVNQTSGBRLSRKLDNNRPVVASLVTVD 1250  
QY 1105 GVHSVTACALVITITDEMTHSITLRLDEMSPERFLSPLGLFIOAVATLATPPDHV 1164

1251 GLHSVTAQCLRVITITTELLANSITVRLENNMOERFLSPLLGRFLEGVAAVATPAEDV 1310  
QY 1165 VFNVOVDTADAGHILANVLSVQPPGCGP--PFLPSEDLOERLYNRSILTASAO 1222  
Db 1311 PIFNIONDIDV--GTYLVNPSALAPRGAGAGAPWFSSEBLOELYRRAALARSLL 1369  
QY 1223 RVLPEDDNLCLEBPCENTNRCVSLRFDSSAPFLASSVLPPIHVPGLRCRCPPGFTG 1282  
Db 1370 DVLPEDDNVCLEBPCENTNRCVSLRFDSSAPFLASSVLPPIHVPGLRCRCPPGFTG 1429  
QY 1283 DYCETEVLDICVSPGPHRCGRSREBGTCLCRDGTGHHGVASASGCTPGVCKNGCT 1342  
Db 1430 DFCETELDLCVSNPCNGSACARRREGTTCVCRPPTGSDCLDTEAGRCVPGVCKNGCT 1489  
QY 1343 CVNLVGFCKDCPCPSGD--FEKPYCOVTTSPFAHSFITERGLRORHFTALSFATKED 1401  
Db 1490 CUDAPNGRCQCPAGANBEGRCVTAASFPSSFWMRGLRQFHLTSLSPATVQGS 1549  
QY 1402 GLLITNGRNEKHDVLALEVOYQVLTFSAGESTTUSPFVPGVSDQMTVOLKYN 1461  
Db 1550 GLLITNGRNEKHDVLALELVAGVRLTYSTGESNTVVSPTVPGGLSDQMTVHLRYN 1609  
QY 1462 KPFLGQTLPOGSEOKVAVVTVVDGDPGVALRPGSVLVNYSCAAQGTGSKSLDITLG 1521  
Db 1610 KERTDALGAGPSPKOKVAVLSVDDCVAVALQFAGLGNYSCAAQVOTSSKSLDITLG 1669  
QY 1522 PLLGVPDLPESPFVRMRQFVGCNRNLQVDSRHLDMADFIANNGTVPCCPAKKVCDN 1581  
Db 1670 PLLGVPDLPESPFVRMRQFVGCNRNLQVDSRHLDMADFIANNGTVPCCPAKKVCDN 1729  
QY 1582 TCHNGTCVNQMDASCBCEPLFGGKSCAQEBANQHFGLSSLVAMH--GLSLPIQPVYL 1640  
Db 1730 PCKNSPFCSEBRSRGSPPVFGKDCQTLTAHBNHFGNGLTSMNFSDMAVSPVYL 1789  
QY 1641 SLMPRTROADGVTLAITRGRSTITLQLEBGMVLSV--BGTGLQASSLLEBRANDGM 1699  
Db 1790 GLAFRTATQVLMQVQAGPSTLQLDRGLSTVTRSG--RASHLLDQVTVSDGM 1848  
QY 1700 HNAQIAL--GASGGRHAL--SPDYGOORABGNLAPRLGHLNITVGGI--PGPAGV 1754  
Db 1849 HILRELEGEPPGRGHVILVSLFSLPQDTMAVSELQGLKVLQHLNGGLRPGSABEA 1908  
QY 1755 ARGFRGCLQAVVSDTPEGVN--LPSHGESLVNVOGCSLPPCCSNPPRANYSYCSMD 1813  
Db 1909 PGLVGCILQVWLGSTPSSPLRPPSH--RNAAPGCVVTAACASGPPRPAKCDLMO 1966  
QY 1814 SYSCSDPGYGDNCNTVNCDLAPCEHOSVCTKRPAPMGYCECPNYLGPYCETRIDP 1873  
Db 1967 TFSCTQPGTYGPGCVDACLNFCQNGSCRLPAPRQYTCDCYGVGHHCHERMDQ 2026  
QY 1874 CPGWMMGHEPTGPCNCDVSKGFPDPCNTKSGECHKENHYRPPGSPPTCLLDCYPTGSL 1933  
Db 2027 CPGWMMGHEPTGPCNCDVSKGFPDPCNTKSGECHKENHYRPPGSPPTCLLDCYPTGSL 2086  
QY 1934 RYCDPEBQCPCKPQVYIGQCRCRCDNPRAYTTCNVNYSDBRAIEMGIMPRRRL 1993  
Db 2087 RSCAPHSQCPKRALGQCNCSDFRAYSAGCRVLYDCPSLNSGVMPPQTFEV 2146  
QY 1994 PAAAPCKSPGTAVHCDERHGMPLPMLFNCSTYTFSELKAPARLQNESGDSRGO 2053  
Db 2147 LATVPCPRALAAARLCEBAGMWLEPDLFNCSTYTFSELKAPARLQNESGDSRGO 2206  
QY 2054 QALLIRNAQHTAGYFSGDVAVVQATRLIAHESSTORGSLATODVHTENILRVGS 2113  
Db 2207 KLAQRLEBVTGHTDYFSQDVAVVQATRLIAHESSTORGSLATODVHTENILRVGS 2266  
QY 2114 ALLDTANKHMLL--QOTEG--GTAMLOHTEAVASALAOHMTYLSPPFIIVPNYI 2169  
Db 2267 ALAPERTGLMALGGRAPGSGSAGVRLHLEBTAATLARKMELTYLAMPGLVTPNLT 2326  
QY 2170 SVRLDKGNF--AGALPRYEA--LRGEOPDLATTVILPESVFRTPPVVAPGGEAO 2225  
Db 2327 SIDRMHPSRPGARRYPRYHNSLFRQDQAMPBHTHVLLPQSPRPSSEVLPJTSSSEN 2386

```

QY 2226 BBEELARRQRHPELSQGEAVASVIIYRTLAGLPHNYDPDKRSIVPRPIINTPVSI 2285
DB 2387 STTSVPPPPAPPEPEPGISIIILVYRTLGILLPAQFOABRRGARNLPQWPNNSVVS 2446
QY 2286 SYHDBELLPRALDKVYVQFRLLEBETKPICVFWMHSILVSGTSGMSARCEVFRN 2345
DB 2447 AVFHGNFRIGLIESISLEFRLQTNANSKICVQWDPGGLABQGVMTADCELVHN 2506
QY 2346 BSHVSCQCHMTSPFALMDSRE--NGEILPLKTLTYALGVTLAAILTFELTLRI 2403
DB 2507 GSHARCRCRTGTGVLMDASPRERLEGULIANTHVAVSVAAVLAAILLSLS 2566
QY 2404 LRSNQHGRNLTALAGLQVLELGINQADLPACTVAILLHFLYCTSPMALIALH 2463
DB 2567 LKSNVGHANVAALGVALLEFLIGIHRTHNQLVCTAVAILLHFLSTFAMLFVQGH 2626
QY 2464 LYRALTEVDVNTGPRFYTGMGVPATITGLANGDEBIGNDPFCLSTYDTLMSF 2523
DB 2627 LYRMQVPEPNVDRGMRFYHALGWMGPAVLGLAVGLDEGYGNDFCWI SYHEPLMSF 2686
QY 2524 AGPVAFAVMSYFLYTIARASCA--AORQGEKGPVSGLOPSPAVILLSTMTLALIS 2582
DB 2687 AGPVVLVIMNGMTFLAARCTSGTQREA--KTSALTILKSSFLILLVLSMTLFLIA 2744
QY 2583 VNSDTLLPYLIPATCNCIGPPIFLSYVLSKVRKALTLAC--SRKPPDPALTTKSTLT 2641
DB 2745 VHSIILAFYIHAAGLCGLQGLAVLLFCYLMNDAARAAWPAICGRKAPEBARPAQGLP 2804
QY 2642 SSYNCPSPADGRQLQ--PYGDSAGLSHSTRSGSKOP-----SYT--PFLIREBA-- 2689
DB 2805 GAVNYTALFEESGGLRITLGASTVSSVSARSGRTOODOSGRSTLRNVIVRSGSAD 2864
QY 2690 -----LNPQGGPPGAG-----DPGSLFLPEGDDQDHPDTSDDLSEDDSGSYASTH 2738
DB 2865 HTDHSLOAHAGTDDVAMFPHDAGA-----DSDDSDLSLEESLSLIPSE 2912
QY 2739 SSSDEEBEERBEAEPGEQMSDILGPAERLPLHSTPKQGGPQKAPVPG----- 2792
DB 2913 SEDNRTGRFQRPICRAAQ-----SERLLTH--PKVDGNDLLSYPALGECBA 2960
QY 2793 -----DROT-----TAKESSGNGAPEERLRKENDALSREGSLGPLGSSAOGHKIL 2839
DB 2961 APCALQTWGSERRIGLDTSKDANNQPPPAL--TSGD-----TSU-----GRQORRKGIL 3011
QY 2840 KKKCLPTISEKSLRLP-----LEOCTGSSRSGSA-----SESGSGGPP 2880
DB 3012 KKR-----LQPLVPTQRCAPELSKRAATLGRRAVPAASYGRIYAGGTSLS 3060
QY 2881 RPPPROSLQEOJLN 2893
DB 3061 QPASRYSRQJLD 3073

```

```

RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34 (1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC cerebellum, olfactory bulb, cerebral cortex, hippocampus and
CC brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domain.
CC -1- SIMILARITY: Contains 8 EGF-like domain.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 Gps domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011528; BAA32459.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR000152; Asx_hydroxy1.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR006210; TRGF_receptor.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PCD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS50221; GPS; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE; PS50227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 31
FT CHAIN 32 313
FT FT 32 2538
FT FT 2539 2559
FT FT 2560 2570
FT TRANSMEM 2571 2591

```

FT	DOMAIN	2592	2599	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2600	2620	3 (POTENTIAL).
FT	DOMAIN	2621	2641	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2642	2662	4 (POTENTIAL).
FT	DOMAIN	2663	2679	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2680	2700	5 (POTENTIAL).
FT	TRANSMEM	2701	2724	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2725	2745	6 (POTENTIAL).
FT	TRANSMEM	2746	2752	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2753	2773	7 (POTENTIAL).
FT	DOMAIN	2774	3313	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	317	424	CADHERIN 1.
FT	DOMAIN	425	536	CADHERIN 2.
FT	DOMAIN	537	642	CADHERIN 3.
FT	DOMAIN	643	747	CADHERIN 4.
FT	DOMAIN	748	849	CADHERIN 5.
FT	DOMAIN	850	952	CADHERIN 6.
FT	DOMAIN	953	1058	CADHERIN 7.
FT	DOMAIN	1059	1160	CADHERIN 8.
FT	DOMAIN	1059	1257	CADHERIN 9.
FT	DOMAIN	1366	1424	EGF-LIKE 1, CALCIUM-BINDING.
FT	DOMAIN	1426	1462	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	1466	1505	EGF-LIKE 3, CALCIUM-BINDING.
FT	DOMAIN	1506	1710	LAMININ G-LIKE 1.
FT	DOMAIN	1713	1749	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	1753	1935	LAMININ G-LIKE 2.
FT	DOMAIN	1937	1972	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1973	2011	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	2012	2044	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	2046	2081	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	2087	2120	LAMININ EGF-LIKE.
FT	DOMAIN	2475	2527	GPS.
FT	DISULFID	1370	1381	BY SIMILARITY.
FT	DISULFID	1375	1412	BY SIMILARITY.
FT	DISULFID	1414	1423	BY SIMILARITY.
FT	DISULFID	1430	1441	BY SIMILARITY.
FT	DISULFID	1435	1450	BY SIMILARITY.
FT	DISULFID	1452	1461	BY SIMILARITY.
FT	DISULFID	1470	1481	BY SIMILARITY.
FT	DISULFID	1475	1491	BY SIMILARITY.
FT	DISULFID	1493	1504	BY SIMILARITY.
FT	DISULFID	1717	1728	BY SIMILARITY.
FT	DISULFID	1722	1737	BY SIMILARITY.
FT	DISULFID	1739	1748	BY SIMILARITY.
FT	DISULFID	1941	1952	BY SIMILARITY.
FT	DISULFID	1946	1961	BY SIMILARITY.
FT	DISULFID	1963	1972	BY SIMILARITY.
FT	DISULFID	1976	1987	BY SIMILARITY.
FT	DISULFID	1981	1999	BY SIMILARITY.
FT	DISULFID	2001	2010	BY SIMILARITY.
FT	DISULFID	2018	2031	BY SIMILARITY.
FT	DISULFID	2033	2043	BY SIMILARITY.
FT	DISULFID	2050	2065	BY SIMILARITY.
FT	DISULFID	2052	2068	BY SIMILARITY.
FT	DISULFID	2070	2080	BY SIMILARITY.
FT	MOD RES	1954	1954	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	838	838	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1173	1173	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1308	1308	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1318	1318	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1640	1640	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1704	1704	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1761	1761	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2173	2173	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2192	2192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2382	2382	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2472	2472	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2504	2504	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	3313	AA; 359348	MM; B11DA09517288764 CRC64;

Qy	Query Match	50.1%; Score 7780.5; DB 1; Length 3313;
Db	Best Local Similarity	53.2%; Pred. No. 0;
Db	Matches 1540; Conservative 414; Mismatches 761; Indels 182; Gaps 44;	
Qy	118 GIGGHLSPGKLTLP-----EEHPCAKAPRLRQSGCKLAQAPL--RAGRRSEBSLG---168	
Db	237 GIGGSLAPAVTAPAPGAPAPESRTAPB--RMRSGFLRGRGFLFRRPGRPFGPTGABA 295	
Qy	169 -----GRRKRNVTAPQFPQPSQATVPENQAGTPVASLRAIDPDGEGRLLEYTM 220	
Db	296 KRLISTNQARSRRAANRRHPQFPQVNYQTLVPENEAAGTAVLRVAQDDPGAGSLVYSL 355	
Qy	221 DALPFRSNQFPGLDPVTGAVTTAEELDREYKSTHVFVYTAODHGMPPRSALATTLVLT 280	
Db	356 AALMMSRSLTEFISIDPGSLRTAAALDRSEMERHLYVTAQDHGSPRLSATTWAVVYA 415	
Qy	281 DPNDDHPVEQEQVEKESIRENLEVEYVLTVAATGDA.PPNNILYRLLESGGSP---336	
Db	416 DKNDAHPVEQEQVEKESIRENLEVEYVLTVAATGDA.PPNNILYRLLESGGSP---GSPAART 471	
Qy	337 --SEYFEIDPRSGVTRTGPVDRBEVESYQLTVEASDQGRDPGRSTTAAYELSYEDND 394	
Db	472 AAAAAFEIDPRSGGLISTGRVDRHEMESYELVEASDQGRGPRSATRVHAIITVLEND 531	
Qy	395 NAPQSEKRYVQVEDVTPGAPVLRVTASDRKGSNAVHYSINSQARQGFYIDAOTG 454	
Db	532 NAPQSEKRYVQVEDVTPGAPVLRVTATDQKANGLVHNIISGSRGFAIDSLTG 591	
Qy	455 ALDVSPIDYETTEKRYTLRVAQDGRPPLASVGLVYVYLDINDNAPIPVSTPQATV 514	
Db	592 EIQVNAIPDPEKERYTALIRQDNGRPLSNITGLASIQQVDINDHSIFVSTFPQSV 651	
Qy	515 LESVPLGYLVAVQALIDADAGNARLEAYRLAGVGHDPPTINNGTMSVAAELDREVD 574	
Db	652 LEMALGHSHVHIQAVDADHGENSLKSLGVASDTEFVINSAGWVSQGLDRESVE 711	
Qy	575 FYSFGEYARHDGTPALTLASASVTVLVNDNNPFTQPEYTVRLNEDAAVGTSTVYTA 634	
Db	712 HYFPEYARHDGSPPLASASVTVLVLDVNDNRPFTKGRYHLRLNEDAAVGTSTVYTA 771	
Qy	635 VDRDASHVTVQITSGNTRNRPSTISQSGGLVSLALPLDYKLEBQYVLAVTASDGTOD 694	
Db	772 VDRDANSALSVQITGNTNRPALSTQSGMGLVTLALPLDYKLEBQYVLAVTASDRLHD 831	
Qy	695 TAOIVVNTDANTHRPVEQSSHYTVNANEDRPACTVVLISATDEDTGENARITYFMEDS 754	
Db	832 HCYVHINITDANTHRPQOSAHYSVSMNEDRPVSGTVVVISASDDVGENARITYLLEND 891	
Qy	755 ITPQRIADDTGAVTQALDYEDQVSYTLATTAHPNGIPQKSDTTYELVNDVNDNAPQ 814	
Db	892 ITPQRIADDTGAVTQALDYEDQVSYTLATTAHPNGIPQKADTTYVEVMVNDVNDNAPQ 951	
Qy	815 PLRDSYQSVYEDVPPFTSVYQISATDRDGLNGVFTFGQDGDGDPFTVESTGIVR 874	
Db	952 PLAGHITGLVSEDAFPFTSVYQISATDRDAHANGVQTFPQNGRGGDDFTIFPISGLVR 1011	
Qy	875 TLRLRDRENVAAQYLRAVAVDKAMPAPATPMEVTVTVLVLDVNDNRPVEQDEDFVFEENS 934	
Db	1012 TVRLRDRENVAAQYLRAVAVDKAMPAPATPMEVTVTVLVLDVNDNRPVEQDEDFVFEENS 1071	
Qy	935 PIGLAVAVTATPDPEGNAQIMQIYVGNIPBYVQGLDIFSGELTALVDLYDRPEEYVL 994	
Db	1072 IYGSVVAQITVADPDGDNAHIMQIYVGNIPBYVQGLDIFSGELTALVDLYDRPEEYVL 1131	
Qy	995 VLOATSAPLVRATVHVVLIDRNDNPPYLGNFELLFNNTYVNRSSSPFGAIGRPAPDP 1054	
Db	1132 VVQATSAPLVRATVHVVLIDRNDNPPYLGNFELLFNNTYVNRSSSPFGAIGRPAPDP 1191	
Qy	1055 DISDSLTSFRRGNELSVLINAAGTEKLKSRALDNNRPLEALNSVLVSDGVSYTAQCA 1114	
Db	1192 DVSDDLFFSFRGNELQVLVYNQTSGEKLKSRKLDNNRPLEALNSVLVTDGHSVTAQCV 1251	
Qy	1115 LKVTITTBMLTHSTLRLBDSMPERFLSPGLGLFQAVAAVTLATPPHHVVVFNQRTD 1174	

```

Db 1252 LKVVITTEELANSLVLENNMWERLSPLEHFEVAVATAPEDVFIFINIDND 1311
Qy 1175 APGGHILNLSVGPDPGPGGP--PPLPSEDIQERLYNRSLLTIAAORVLPDNDIC 1232
Db 1312 V-GGTVLANSFSLAPRAGAGAGAPWSEBEOEOLYVRALAAARSLDLVPDNDVC 1370
Qy 1233 LREPCENYKCVSLAFDSSAPFIASSVLFRPIHPVGLRCRCPGFTGDYCETEDLC 1292
Db 1371 LREPCENYKCVSLAFDSSAPFIASSVLFRPIHPVGLRCRCPGFTGDYCETEDLC 1430
Qy 1293 YSRPCHGCRSREGYTCLCRDGYGHECEVSASGRCTPVCNKGCTCVLLVGGRR 1352
Db 1431 YSRPCHGCRSREGYTCLCRDGYGHECEVSASGRCTPVCNKGCTCVLLVGGRR 1490
Qy 1353 CDPCSGD-PEKPYCOVTTSRFPNHSFTTFRGLRORFHTLSPATKEDGILLNGREF 1411
Db 1491 CQCPAGAGFEGPCFVAAARSPSPSPFMFRGLAQRFHTLSTLSFATVQBSGLLFTNRLN 1550
Qy 1412 EKRDFAALEYIOBOVOLTFSSAGSTTVSPFVGVSDQGMATVOLKYTNKPLAQGTLP 1471
Db 1551 EKRDFAALEYIOBOVOLTFSSAGSTTVSPFVGVSDQGMATVOLKYTNKPLAQGTLP 1610
Qy 1472 QGSEKQVNAVTVDCDYGVALRGSVLGNYSCAAGTGGSKSLDLTGPLLLGVPDL 1531
Db 1611 QGSEKQVNAVTVDCDYGVALRGSVLGNYSCAAGTGGSKSLDLTGPLLLGVPDL 1670
Qy 1532 PESFPVRMRQFVGCNENLQVDSRHIDMADFIANNGTVPGCCPAKNVCDSTNCGTGVN 1591
Db 1671 PESFPVRMRQFVGCNENLQVDSRHIDMADFIANNGTVPGCCPAKNVCDSTNCGTGVN 1730
Qy 1592 QMDAFSECEPLRGFGKSCAQMANNPOHFGSSLVAN-HGLSLDIPQWYLSLMTFRQAD 1650
Db 1731 RMGFGFCDCDVGKXDCRLTMHPYHFGNGTSLMDFGNDMVSYPWYLGSLFRTARK 1790
Qy 1651 GYLLQAITRGRSITTLQIBRGHMLSVETGLOASSLRLEPGANGDWTHAOL--GA 1708
Db 1791 GYLLQAITRGRSITTLQIBRGHMLSVETGLOASSLRLEPGANGDWTHAOL--GA 1850
Qy 1709 SGGRGHAII--LSFDYQOQRAEGNIAPRLHGLSLNITVGGIP-----GPAGGVARGFRG 1760
Db 1851 SGGRGHAII--LSFDYQOQRAEGNIAPRLHGLSLNITVGGIP-----GPAGGVARGFRG 1905
Qy 1761 CLOGVAVSDTPEBVNSL-DSHGESINVEQCSLPDPCSNPCPNANSYCSNDYSYSCC 1819
Db 1906 CLOGVAVSDTPEBVNSL-DSHGESINVEQCSLPDPCSNPCPNANSYCSNDYSYSCC 1963
Qy 1820 DPGYVGNCTNVVDLNPCEHOSVCTRKSAFHGYTCEBPNTYGFYCCERRIDPCRRGMW 1879
Db 1964 DPGYVGNCTNVVDLNPCEHOSVCTRKSAFHGYTCEBPNTYGFYCCERRIDPCRRGMW 2023
Qy 1880 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPGSPYCLLCYPTGSLRYCDE 1939
Db 2024 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPGSPYCLLCYPTGSLRYCDE 2083
Qy 1940 DGQCCPKPVYIGQCRCNDNPAEVTNNGCEVYDSCPAIEAGIWMPTRFGLPAAAPC 1999
Db 2084 DGQCCPKPVYIGQCRCNDNPAEVTNNGCEVYDSCPAIEAGIWMPTRFGLPAAAPC 2143
Qy 2000 PKGSF-----GTAVRHODEHGRGLPRLFNCTSTTSELKGFARLQRMESGLDSRSQO 2054
Db 2144 PKGSF-----GTAVRHODEHGRGLPRLFNCTSTTSELKGFARLQRMESGLDSRSQO 2203
Qy 2055 LALLLNATQHTAGYSGSDVKAVALATRLLAHESITORGFSATODVHTENTILVGA 2114
Db 2204 LALLLNATQHTAGYSGSDVKAVALATRLLAHESITORGFSATODVHTENTILVGA 2263
Qy 2115 LLDTPANKRMELI--QOTEG--GTAMLIQYEAIVASALAKMNHITLSPYITENIVIS 2170
Db 2264 LLDTPANKRMELI--QOTEG--GTAMLIQYEAIVASALAKMNHITLSPYITENIVIS 2323
Qy 2171 VVALDK-GNFAGA-KLPRYA--LRGQRPDLTETVILPESVREPRPVVR----- 2218

```

```

Db 2324 IDRMQPSSSOGAHRYPRYHSLFRGODAMDHTVLLPSSQSPSPSEVLPTSSNAENA 2383
Qy 2219 AGPGEAOBEELARORRHPBELSGEAVASVITYRTIGLPHANDPDGRSLRVPRPII 2278
Db 2384 TASGVSPAPAL-----EPSEBPGISTVILLVYALGGLLAQOAEKRGARLPONPW 2437
Qy 2279 NTPVVISVHDBELLPRALDKPTVQFALLTEBERIKPICVFMNHSILVSGTGVNSARG 2338
Db 2438 NTPVVISVHDBELLPRALDKPTVQFALLTEBERIKPICVFMNHSILVSGTGVNSARG 2497
Qy 2339 CEVFRNBSHVSQCCNHTSPVLDVSRRE--NEELPLKTLTYVALGVTALAILTFP 2396
Db 2498 CEVFRNBSHVSQCCNHTSPVLDVSRRE--NEELPLKTLTYVALGVTALAILTFP 2557
Qy 2397 PTLRLIRSNONGIRRMUTALALAOVFLIGNOADLPACTYATILHPLYCTPSW 2456
Db 2558 PTLRLIRSNONGIRRMUTALALAOVFLIGNOADLPACTYATILHPLYCTPSW 2617
Qy 2457 ALLEALHLYRALTEVRDVNTGPMRFYMLGMVPAFITGLAVGLDPEGYNDPCWLSIY 2516
Db 2618 ALLEALHLYRALTEVRDVNTGPMRFYMLGMVPAFITGLAVGLDPEGYNDPCWLSIY 2677
Qy 2517 DTLVSPFAPVAFVMSVFLYITLAARASCAARQGFKEGPFVSGLOSPFVALLSATW 2576
Db 2678 DTLVSPFAPVAFVMSVFLYITLAARASCAARQGFKEGPFVSGLOSPFVALLSATW 2737
Qy 2577 ILALISVNSDTLLFHYLPATNCICQGPPIFLSYVLSKVRKALALAC--SRKSPDPALT 2635
Db 2738 ILALISVNSDTLLFHYLPATNCICQGPPIFLSYVLSKVRKALALAC--SRKSPDPALT 2797
Qy 2636 TKSITLSTSYNCPSPYADRLYO-PYGDASGLSHSTRSGKQP-----SYI--PFLR 2685
Db 2798 TKSITLSTSYNCPSPYADRLYO-PYGDASGLSHSTRSGKQP-----SYI--PFLR 2857
Qy 2686 E-----ESALNPGQCPGLG-----DPGSLFLEGDDQDHPPTDSDSLSDDDSG 2732
Db 2858 E-----ESALNPGQCPGLG-----DPGSLFLEGDDQDHPPTDSDSLSDDDSG 2905
Qy 2733 SYASTHSDSEEBEEREEBEAFAPEQGWDSILGPARLPLHSTPKQG-----PGG 2786
Db 2906 SYASTHSDSEEBEEREEBEAFAPEQGWDSILGPARLPLHSTPKQG-----PGG 2955
Qy 2787 K--AP-----WPGD--FGTA--KESGNGAPEERLRENGDALSHRGLGAPLGSSAOPH 2835
Db 2956 K--AP-----WPGD--FGTA--KESGNGAPEERLRENGDALSHRGLGAPLGSSAOPH 3006
Qy 2836 KGIILKKKCLPTISSEKSLRLP-----LECTSSSKGSSA-----SEGRG 2876
Db 3007 KGIILKKKCLPTISSEKSLRLP-----LECTSSSKGSSA-----SEGRG 3055
Qy 2877 GPPRPPROSLQOQLN 2893
Db 3056 GPPRPPROSLQOQLN 3072

```

```

RESULT 8
CLR3 MOUSE
ID CLR3 MOUSE STANDARD; PRT; 3301 AA.
AC 091210; Q9RS0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caderlin BGF LAG seven-pass G-type receptor 3 precursor.
GN CELSR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Butelcostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBT_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.
RC STRAIN=C57BL/6;
RX MEDLINE=2183955; PubMed=11850187;
RA Tisbet F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;

```





FT DISULFID 1974 1985 BY SIMILARITY.  
 FT DISULFID 1979 1997 BY SIMILARITY.  
 FT DISULFID 1999 2008 BY SIMILARITY.  
 FT DISULFID 2016 2029 BY SIMILARITY.  
 FT DISULFID 2031 2041 BY SIMILARITY.  
 FT DISULFID 2048 2063 BY SIMILARITY.  
 FT DISULFID 2050 2066 BY SIMILARITY.  
 FT DISULFID 2068 2078 BY SIMILARITY.  
 FT MOD\_RES 1952 1952 BY SIMILARITY.  
 FT CARBOHYD 623 623 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 2713 2713 L -> LR (IN REF. 2).  
 FT CONFLICT 3024 3024 R -> P (IN REF. 2).  
 SQ SEQUENCE 3501 AA; 358455 MW; A6B18P2DF7F4DEB6 CRC64;

Query Match 49.7%; Score 7732.5; DB 1; Length 3301;  
 Best Local Similarity 53.5%; Pred. No. 0;  
 Matches 152; Conservative 411; Mismatches 734; Indels 181; Gaps 44;

QY 157 RAGERSEESLG-----GRRKNVNTAPQPPSYQATVPENQAGTFVASTIRA 205  
 DB 281 RPPGRPPGFPFGPEAKQILSTNQARPRRAARHPQFPQYNYQTLVPENBAAGTSVLKVA 340  
 QY 206 IDPEGAGLELYMDLAFDSRNOFPSLDPVNGAVNTAEBDBETSTVFRRTADHG 265  
 DB 341 QDPGPGAGRLTSLALMNSRSLPSPIDQSLITPAALDESEMERHYLRATQDDHG 400  
 QY 266 MPRRSALATLTITVTDNDHPVEEQEYKESLENTLEVGVLTVAATDGDAPNNIL 325  
 DB 401 SPRLSATWVAAYVADNDHAPVEQAOYRETLRENTBEYGPITQLAATGADAPNANLR 460  
 QY 326 YRLLEGSGSP-----SEVPEIDPSGVITRGPVDRKEEVSYQLTVEASDQGRDGP 379  
 DB 461 YRFV---GSPAVRTAAALAAFEIDPRSGLISTSGRVDRHMESEYELVEASDQGRDGP 516  
 QY 380 STTAAVFLSEVDNDNAPOSEKRYVVOVREDYTGAPVLRVNASDQKSNVAKSIM 439  
 DB 517 SATVRVHITVLDENDNAPQEGEKRYVAQVREDVAPHTVVLRTVATDKDANGLVHYNII 576  
 QY 440 SGNARGFYLDACOTGALDVVSPLDYETTKETYLVRRAODGGRPPLSVSGLVTVQVLDIN 499  
 DB 577 SGNSRGHFAIDSLTGEIOWNAPLDFEABRYALRIKQADGRPPLSNTGLASIQVDIN 636  
 QY 500 DNAPIFVSTFPQATVESVPLGYIVLHVQALDADAGNABLYRLAGVGHDPFTINGT 559  
 DB 637 DHAPIFVSTFPQVSLNAPELGSHVIHIQAVDADHGENSLREYSLTVAADTDFVINSAT 696  
 QY 560 GWSVAAELDRERDVPYSGVEARDHCTPALTASASVTVLVNKNPFTFOPEYTVL 619  
 DB 697 GWSVSGPLRBSSEVHEFPVEARDHSGPPLSASASVTVLVNKNPFTFOPEYTVL 756  
 QY 620 NEDAAVGSVTVTASAVDRDASHVTVYQITSGTRNRPSITSGSGGLVSLALPLDYLER 679  
 DB 757 NEDAAVGSVTVTASAVDRDASHVTVYQITSGTRNRPSITSGSGGLVSLALPLDYLER 816  
 QY 680 QYLAIVTASGCTODTAQIVVNTDANTHRPFQSSHVTVVNVNEDRPAQTIVVLISATDE 739  
 DB 817 YFXLVLTASDRALHDHCYVHINITDANTHRPFQSAHVSVMNEDRPAQSVTVVIVIASDD 876  
 QY 740 DTGENARITYFMEDSIPOFRIDADTGAVTQAELDYEDQVSYTLATITARNGLPOKSDPT 799

DB 877 DVGENARITYLLEEDNLEQFRIDADSGAITLAQPLDYEDQVYTLATITARNGLPOKADTT 936  
 QY 800 YLELVVDVNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLANGVFTYFOGDD 859  
 DB 937 YVEVAVDNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLANGVFTYFOGDD 996  
 QY 860 GDDGPIVESTGIVRTLRRLRENVAAQVLAAYADKMPARTPMEVTVLVLDVNNPP 919  
 DB 997 GDDGPIVESTGIVRTLRRLRENVAAQVLAAYADKMPARTPMEVTVLVLDVNNPP 1056  
 QY 920 VFEODEFDVVEENSPIGLAVANATADPDGNTAQIMYOIVEGNIPVEFOLDIFSGELT 979  
 DB 1057 VFEODEFDVVEENSPIGLAVANATADPDGNTAQIMYOIVEGNIPVEFOLDIFSGELT 1116  
 QY 980 ALVDLDEDEPEVLYVQATSAPLVSATVHVLVLRNDNPPPLGNETLFPNNYTRSS 1039  
 DB 1117 ALVDLDEDEPEVLYVQATSAPLVSATVHVLVLRNDNPPPLGNETLFPNNYTRSS 1176  
 QY 1040 SPFGAIGRVAHPDIDSLTYSFERGNETSLVLNASTGELKLSRALDNNRPLEAING 1099  
 DB 1177 TFGSGIIGRIPADPDVSOHLFYSFERGNETSLVLNASTGELKLSRALDNNRPLEAING 1236  
 QY 1100 VLVSDGVSYTAQCALRVITITDEMLTSTTLRLDMSPERFLISPLGLEIQAUAATLAT 1159  
 DB 1237 VLVSDGVSYTAQCALRVITITDEMLTSTTLRLDMSPERFLISPLGLEIQAUAATLAT 1296  
 QY 1160 PPDHVVVFNQRTDAPGHIILNLSLVQPPPPGPP--PFLPSDQRLVANSLSLT 1217  
 DB 1297 PPDHVVVFNQRTDAPGHIILNLSLVQPPPPGPP--PFLPSDQRLVANSLSLT 1355  
 QY 1218 ALSAORVLPDDNIDCLREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHPVGLRCRCP 1277  
 DB 1356 ALSAORVLPDDNIDCLREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHPVGLRCRCP 1415  
 QY 1278 PGTGTGYCTEVDLCTSRPCGPHGCRSRBGYTCICRDGYTGEHCEVASARSCTPVGVC 1337  
 DB 1416 PGTGTGTCTEVDLCTSRPCGPHGCRSRBGYTCICRDGYTGEHCEVASARSCTPVGVC 1473  
 QY 1338 KNGTCVNLVGGFKCDKPSGD--PEKPYCOVTRSPRSPHSTIFRGLRORFHTLLASFA 1396  
 DB 1474 KNGTCVNLVGGFKCDKPSGD--PEKPYCOVTRSPRSPHSTIFRGLRORFHTLLASFA 1533  
 QY 1397 TKERDGLLYNGRFRNKDHFVALLVIOVOLTSPSAGSTTVSPVPGSVSDGMHTVQ 1456  
 DB 1534 TKERDGLLYNGRFRNKDHFVALLVIOVOLTSPSAGSTTVSPVPGSVSDGMHTVQ 1593  
 QY 1457 LKTYNKPGLQGTGLPGGSEBOKYAVVTVDCDTGVALRFGSVLGNYSCAAQGTGSSKXS 1516  
 DB 1594 LKTYNKPGLQGTGLPGGSEBOKYAVVTVDCDTGVALRFGSVLGNYSCAAQGTGSSKXS 1653  
 QY 1517 LDTGPIILGGVDDPESPFPVMBROFVGMGRNLOVDSRHIDMADFLANNGTVPGCCAKXN 1576  
 DB 1654 LDTGPIILGGVDDPESPFPVMBROFVGMGRNLOVDSRHIDMADFLANNGTVPGCCAKXN 1713  
 QY 1577 VCDSTNCHNGGTCVNMOMAFSCBPCPLGFGGKSCAOMANPOHFLGSSLVAM--HGLSLPTS 1635  
 DB 1714 VCDSTNCHNGGTCVNMOMAFSCBPCPLGFGGKSCAOMANPOHFLGSSLVAM--HGLSLPTS 1773  
 QY 1636 QPWTLSIMFRTOADGVLLQA-----ITGRSTITTLQLEBHVMLSVBGTGLQA 1684  
 DB 1774 QPWTLSIMFRTOADGVLLQA-----ITGRSTITTLQLEBHVMLSVBGTGLQA 1823  
 QY 1685 SSIIRLEPGRANGDMWHAQAL--GASGGPGHAI--LSFDYQOQABAGLGRILGHILS 1740  
 DB 1824 SSIIRLEPGRANGDMWHAQAL--GASGGPGHAI--LSFDYQOQABAGLGRILGHILS 1882  
 QY 1741 NITVGGIPGPAGVA--RGRFGGLQGVVSDTBEVNS--LDPHSGSINVEQCSLPDPCD 1798  
 DB 1883 NITVGGIPGPAGVA--RGRFGGLQGVVSDTBEVNS--LDPHSGSINVEQCSLPDPCD 1940  
 QY 1799 SNPCPANSYCSNDMSYSCSDPGYTGDNCTVNCNCPCHDSVCTKRKSAHGTTCCEP 1858

```

Db 1941 SGPCPBHADCDLWQTFSCRCRPGYGGCVADACLINPCQNGSCRLQGAHPGYTCDCV 2000
Qy 1859 PNVLAPYCEIRIDPCPRGMWGHPTGFCNCNDVSKGPPDNCIKTSGBCHKENHYRPGS 1918
Db 2001 SGYFGQHCHHHVDQCCPRGMWSPFCGFCNCNDVSKGPPDNCIKTSGBCHKENHYRPGS 2060
Qy 1919 PTCLICDCYPTGSLSRVCDPEDGQCPKPGVYIGRQCDRCNDPFAVTTNGCEVNYDSCPR 1978
Db 2061 DSCILPCDCYPTGSLSRVCDPEDGQCPKPGVYIGRQCDRCNDPFAVTTNGCEVNYDSCPR 2120
Qy 1979 ALEAGIWMPTRTFGLPALAPCKSGRGTAVHSCBHRWMLPBNLFNCSTIFSEIKGFAE 2038
Db 2121 SLRSGVWMPQTRFGLATVPCCRGALGAVALRCBDDQGLBDFLNCSTIFSEIKGFAE 2180
Qy 2039 RLQRNSELDSGRSQQLALLLNATOTAGYFGSDVKAVALATRLLAHESRQFGLSA 2098
Db 2218 GLEINKTLDVTEAKKLAQRLREVTGCDHDFSDQVRYTALUAVILAFESHQCGFGTA 2240
Qy 2099 TQDVHFTENLLRVSGALLDPTANKRWELI-QQTBG--GTAWLQHYEAVASALAQNRH 2154
Db 2241 TQDAHFENELMWAGSALLAPETGHLMAALQGRAPGSGSAGLVQHLEBYAATLARWEL 2300
Qy 2155 TYSPLFTVTNIVSVVRLD-KGNFAGA-KLPRYEA--LAGEOPDELTYYVILPESVYR 2210
Db 2301 TYLNVGLVTPVIMLSIDRMHPSTQCARRRYPRYHMLFRGQDAMDHTTVLILPSQASQ 2360
Qy 2211 ETPPVVRPAGPG--EAOEPEELARRQRHPELSQGEAVASYIYRTLAGLLPHNYDPKR 2268
Db 2361 PPSFVLTSSNAENAKATASVSPAPLEPSEBPISIVILLRYALGGLLPAPQOMARR 2420
Qy 2269 SLRVPKRIINTPVVISVHDBELLPALDKPTVQVQRLITEERTYPCVPMNHSILV 2338
Db 2421 GARLPONVWNSPVVSVAVFGRNFLRVLVSPIMLEFRLLQTNARSKAICVQMDPPGPT 2480
Qy 2329 SGTGMSARGGEVAFRNSHVSQCCNMTSPFALMDVRRR--NGEILPLKTLTPYVALGV 2386
Db 2481 DQHGKMTARDEBELVHRNSHARCSCRGTRGVLDMSPRELBEDLAVFTHVAV 2540
Qy 2387 TLAAALLTFPFLTLRLIRSNQHGIRNMLTAALGLAOLVPLGGINQADLPRACTVIAL 2446
Db 2541 SVTLALVLAVALILSRSLKSNVREGIHANVAALAGVABELFLGIRHTNQOLCTRAVAL 2600
Qy 2447 HPLVICTSSNALLKHTYRALTERTVDVNTGPMRTYMLGNGVPAFTGLAVGLDPBXYG 2506
Db 2601 HYFPLSTFAMLLVQGLHYRMOVPRANDRGMERYHLAGVPAVLLGLVGLDPBXYG 2660
Qy 2507 NPDFCMLSIYDTLISFAPGVAFAVMSVFLYTLAARASCA-AQOGPKKKGPVSGLOPS 2565
Db 2661 NPDFCMLSIHBLISFAPGPIVLYVWNGTMEFLAARTSCSTGQREA--KTSVILTRSS 2718
Qy 2566 PAVILLLSATWMLALLSVNSDTLTFHYLPATCNCIQGPFIPLSYVLSKEVRKALKAC- 2624
Db 2719 FLILLVLSASWLFGLAVNHSILAFHYLAGLGLGLAVLILFVLANADARAATPACL 2778
Qy 2625 SRKPSDDALTTKSTLTSYNSCPSPYADGRLYQ-PYGSASGSIHSTSGSASQP----- 2677
Db 2779 GKKAAPBETTRAPAPGPGSAYNTALFESGILIRITLGLASTVSVSASRSGAQQDQSGRG 2838
Qy 2678 -SYI--PLIRB-----ESALNPGQGPPIG-----DPGSLILBGOQDQHPDPTSD 2721
Db 2839 RSYLRDNLVLRHGSTAHTERSLOAHAGPTLDVAMFRDACA-----DSISD 2886
Qy 2722 SDSLIEDDQSGSVASTHSSESEEBEERAEAFPGQWMSLLPGARERPLHSTPDXG 2781
Db 2887 SDSLIEERSLISIPSESEEDNGRTGRFORPLRAAQ-----SERLLAHPKQVDG 2936
Qy 2782 G-----GPGK--AP-----WPGD--FGTTA--BESSGNAPREERLENDALSRBSL 2824
Db 2937 NDLSIYWPALGECRAPCALQMGSEBRRLGDSNDANNNQPELAL-TSGD-----ERSL 2991
Qy 2825 GPLPSSAQPHKILKKKCLPTISEKSLRLPLTEQCTGSSRGS-----SASRGSRG 2877
Db 2992 -----GRAGRGRKGIKRN-----LOYPL--VPGRGRTBELSWCAATIGHAV 3033

```

```

Qy 2878 PP-----PREPPROSLQEOLN 2893
Db 3034 PAASYGRIVAGGCGSLGSPASRYSSREQD 3064

RESULT 9
STAN DROME
ID STAN DROME STANDARD; PRT; 3579 AA.
AC 09V5N8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE protocadherin-like wing polarity protein stan precursor (Starry night
protein) (Flamingo protein).
GN STAN OR FMI OR CGI1895.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NX NCBI_Taxid=7227;
[1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=Embryo;
RX MEDLINE=20025940; Pubmed=10556066;
RA Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J.,
RA Adler P.N., Park W.J.;
RT "The Drosophila tissue polarity gene starry night encodes a member of
RT the protocadherin family."
RL Development 126:5421-5429(1999).
[2]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=99418630; Pubmed=10490098;
RA Ueki T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of Fz/Dishevelled."
RL Cell 98:585-595(1999).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolintsov S.,
RA Bokoyva D., Botchan M.R., Bouck J., Broksstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J.B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemisun J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,

```

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Massarman D.A., Weinlock G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Wocley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP REVISIONS.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Mera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu I., Berman B.P.,  
 RA Bellen Court B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.,  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 CC -1- FUNCTION: Involved in the fz signaling pathway that controls wing  
 CC tissue polarity. Also mediates homophilic cell adhesion. May play  
 CC a role in initiating prehair morphogenesis. May play a critical  
 CC role in tissue polarity and in formation of normal dendrite  
 CC fields.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: In the pupal wing, expressed at relatively  
 CC even levels in all regions. Abundant in 6-9 hour embryos.  
 CC Expressed at higher levels in pupae than larvae.  
 CC -1- DEVELOPMENTAL STAGE: At 12 hours after puparium formation (apf),  
 CC expressed evenly at cell boundaries. By 30 hours apf, expression  
 CC is concentrated at proximal and distal cell boundaries with little  
 CC or no expression at anterior and posterior boundaries. When  
 CC prehairsts emerge at 30-36 hours apf, expression becomes evenly  
 CC distributed again along the whole cell boundary.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: Contains 8 cadherin domains.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF172329; AAF02618.1; -;  
 DR EMBL; AB028498; BAA84069.1; -;  
 DR EMBL; AB003828; AAF58763.3; -;  
 DR HSSP; P08709; 1BP9;  
 DR FLYBASE; FBgn0024836; atan.  
 DR GO; GO:0016021; C:integral to membrane; NMS.  
 DR GO; GO:0008014; C:calcium-dependent cell adhesion molecule ac. .; IMP.  
 DR GO; GO:0005057; F:receptor signaling protein activity; IMP.  
 DR GO; GO:0016358; P:dendrite morphogenesis; IMP.  
 DR GO; GO:0007222; P:fizzled receptor signaling pathway; IMP.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 DR InterPro; IPRO002126; Cadherin.  
 DR InterPro; IPRO00742; EGF 2.  
 DR InterPro; IPRO01881; EGF Ca.  
 DR InterPro; IPRO06209; GPCR like.  
 DR InterPro; IPRO06832; GPCR\_secretin.  
 DR InterPro; IPRO01879; hormn\_receptor.  
 DR InterPro; IPRO02049; laminin\_EGF.  
 DR InterPro; IPRO01791; laminin\_G.  
 DR InterPro; IPRO00203; PKD\_cys\_rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.

DR Pfam; PF00008; EGF; 3.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00053; laminin\_EGF; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 8.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00282; Lamm; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 6.  
 DR PROSITE; PS00268; CADHERIN\_2; 8.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 3.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECP\_F2\_4; 1.  
 DR PROSITE; PS00025; LAM G DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR PROSITE; PS00221; GPS; 1.  
 KW Cell adhesion; Developmental protein; G-protein coupled receptor;  
 KW Calcium-binding; Repeat; Signal; Transmembrane; EGF-like domain;  
 KW Laminin EGF-like domain; Glycoprotein.  
 FT CHAIN 1 3579  
 FT FT 30 3579  
 FT DOMAIN 30 2816  
 FT TRANSMEM 2817 2837  
 FT DOMAIN 2838 2845  
 FT TRANSMEM 2846 2866  
 FT DOMAIN 2867 2883  
 FT TRANSMEM 2884 2904  
 FT TRANSMEM 2905 2919  
 FT TRANSMEM 2920 2940  
 FT TRANSMEM 2941 2959  
 FT TRANSMEM 2960 2980  
 FT TRANSMEM 2981 3000  
 FT TRANSMEM 3001 3021  
 FT TRANSMEM 3022 3031  
 FT TRANSMEM 3032 3052  
 FT TRANSMEM 3053 3579  
 FT DOMAIN 360 464  
 FT TRANSMEM 465 581  
 FT TRANSMEM 582 689  
 FT TRANSMEM 690 794  
 FT TRANSMEM 795 897  
 FT TRANSMEM 898 1007  
 FT TRANSMEM 1008 1113  
 FT TRANSMEM 1114 1220  
 FT TRANSMEM 1220 1538  
 FT TRANSMEM 1538 1753  
 FT TRANSMEM 1756 1792  
 FT TRANSMEM 1795 1963  
 FT TRANSMEM 1965 2000  
 FT TRANSMEM 2091 2126  
 FT TRANSMEM 2744 2802  
 FT TRANSMEM 140 143  
 FT TRANSMEM 140 159  
 FT TRANSMEM 155 159  
 FT TRANSMEM 2567 2579  
 Query Match 31.0%; Score 4817.5; DB 1; Length 3579;  
 Best local similarity 34.9%; Pred. No. 6; 8e-223;  
 Matches 1108; Conservative 461; Mismatches 1005; Indels 599; Gaps 71;  
 QY 171 RRRNVTAPOPFPSPYOATVPENOPAGTPVASTAIRAIDPDGEAGRLTYTMDALFDSRSNQ 230  
 Db 349 RELRNQSPYFPOALYVASVLEBPQAGAAVTVRRARDPDS---VVYSVSLDERSQS 405  
 QY 231 FSLDPVTCNAVTAELDRSTSTVPRVTADQHGMPRSALATLTILYTDNRDHPVE 290



```

QY 2265 -PDKSLRVKRPILNT----- 2280
DB 2552 EOEYRLEIEBKLPSSSSSSSSSGTEQOQVVEVDVPRAPTSSEQOIEDIRITAHB 2611
QY 2281 ---FVSI-----SVHDEELP----- 2295
DB 2612 IPPVSSVEQOEAASDDBGERERHINLDDIFHNGSGEVI SPSPENLNTYGVIS 2671
QY 2296 -----RALDKPTVQF---RLLETER----- 2313
DB 2672 STGSDQPKGENEAVYEDRLVYKQVEITYPSEQMOQEQVYVNSLQSPHLAQIKLQMW 2731
QY 2314 -----RTKICVFNHNSILVSGIGWSAGCEV-----VFNESHVSCQ 2352
DB 2732 LDVDSARFGRPSNQCVRWN-----SFTNOMTRIGCCYTERIPDFGDNPAQAQALIVNCS 2786
QY 2353 CNHNTSPAVLMDVSRRENGEILPLKLT-----TYVALGVTLALLTLFFFLTLRLRSN 2407
DB 2787 CTHISSYAVIVDVDPED---IPERSLVQVTSYSAFLVSLPLILGYLALALIRGOQTN 2843
QY 2408 QHGIRRLTALGLAQVFLGIGNO---ADLPACTVIALHFLYLCTPSWALTBALH 2463
DB 2844 SNTIHQNI VLVCFCAELLFVGMSRRQLSEFPCKLTALCLHYFWLAFAWTTVDVCH 2903
QY 2464 LYRALTEVDVNTGPMFYTMLGWSVAFITGLAVGIDPREGYNDPCMTSTYTLTWSF 2523
DB 2904 LYRLTLTRDININHPGPFYPMGGAIPAIVGSLVGRABEYGNLSLCTMSTVEPVMWL 2963
QY 2524 AGPAPAVMSVFLYIILAAASCAARQGFQFQKPGVSG---LQSPFVALLLSATV 2576
DB 2964 VGPILAGSVNLLIFVSVDA-----FTLKQHVLFQGLKRLTLMLSVSLPLMGVMW 3016
QY 2577 LIALLSVNSPTLLFHYLPATCNCIQGPFILSYVLSKEVAKALKACSR----- 2626
DB 3017 VLAIVLAASEHQSLLSLLSGVVLLHAFCLIGCIITINKRARENLOKRCRCMGKRVLLD 3076
QY 2627 -----KPS-----PDPALTKSTLTSSVNCSPFADG 2653
DB 3077 SSNWSVSSNNVNAAPNSFLASGYTTTRNIGISASSSTSTSTKTS---SSPYSDG 3133
QY 2654 RLYOPYDASAGSLHSTRSGKQPSYIPLILREBSALNPGQPGGLDPSLFLGQDOQ 2713
DB 3134 QLRQT-----STSTSNVNSASDAPSLRGESESTTGRSGEKEP-----SKROR 3178
QY 2714 HDEPTDSLSLSDDOGSSTAGTHSSDSEEBE-----EEEE 2751
DB 3179 KDSOSGET-----DGRSLKSLASHSSDDDESRATSGSTRSTAVSTPAVLNITEHV 3233
QY 2752 EAPAPGE-----QGDLSLGP-----GAEKLP---LHSTPKDGGGPGKAPMPDPRGT 2796
DB 3234 QATTPPLNVVQSPQLPSPVAKPYAPARWSSQLPDAYLQSPPNIG---RMSQPTGS 3286
QY 2797 TAKSSNGAPBERTLRNGDALSRGSLGPLPGSSAQPHKGLKKCLPTISEKSL--- 2853
DB 3287 DNEHVHQA-----KMTISPNLPLN-----PULTTSTYLOQH 3318
QY 2854 ---LRLP---LEOCTSSKSSASBSGSGRGP---RPPROSLQEOQLNG 2894
DB 3319 HNKINMPISILNIDARBEYEDSLVGRGEYDPDKYSGYXPPHVGSEKDYGP 3371

```

```

CC Ephydroidea; Drosophilidae; Drosophila.
CN NCBI_TaxID=7227;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92069752; PubMed=1959133;
RA Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
RA Goodman C.S.;
RT "The fat tumor suppressor gene in Drosophila encodes a novel member
RT of the cadherin gene superfamily.";
RL Cell 67:853-868(1991).
RN RP SEQUENCE FROM N.A.
KC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ainsworth K., Breen J., Chen L., Culbertson J.M., Drenth J.,
RA Eickbush S.H., Gish N.B., Grigoriadis A.A., Heath W.E.,
RA Johnson D.L., Lasker J.M., Leighton J.L., Lin X., Little J.,
RA Marra M.A., McIninch J., Miller L., Muzny D.M., Nelson D.L.,
RA Penhallow G., Rozen S., Saksela E., Sander C., Schaefer F.,
RA Shinnar R., Sussman M., Telford I., Umayahara T., White O.,
RA Yee B.C., Zeng L., Zheng X.H., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN. ACTS AS A
CC TUMOR SUPPRESSOR. REQUIRED FOR CORRECT MORPHOGENESIS.
CC -1- SIMILARITY: Contains 34 cadherin domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-bb.ch).
CC
DR EMBL; MG0537; AA28530.1; -
DR EMBL; AE003577; AAP51036.1; -
DR HSSP; P00740; 1EDM.
DR FLYBase; Fgn0001075; ft.
DR GO; GO:0005887; C:intracellular plasma membrane; NAS.
DR GO; GO:0008014; P:calcium-dependent cell adhesion molecule ac. .; NAS.
DR GO; GO:0016339; P:calcium-dependent cell-cell adhesion; NAS.

```





```

QY 428 KGSNAVHYSTIMSGNARQFYIDAO-----GALD--- 457
Db 2635 DGENAVVHTI---SGRDQHFYDINTKXGVSTKLELTKTKSHDDLTYTIVISAMDGE 2681
QY 458 -----VSP----- 461
Db 2682 QSLSKELVILIRPELPTPTAYMANSHFAMSEBVRCKMTTKSATSPKGLVGRY 2741
QY 462 -----IDYETKEVTLVRADQDGRPPLSVNGLVYQV 495
Db 2742 AIAAGIMGDLRVPNSGLISVGQGLYELTHLYEIMIBADGOTPELSRVT-LITLNV 2800
QY 496 LDINDNAPIFVSTPQATVLESVPGLYVLAHQADADAGNARLETRLAGVGHDFPTI 555
Db 2801 TDADNAPVEMQOLYNAEVLBEESPQOLIAVVKASDRSGNGVYTLQ---NDPQDTP 2857
QY 556 N-NGTGMISVAALDREBVDYFSGVEARDGTPALTASASVTVLVDVNNPPTFOPE 614
Db 2858 BITSGELIYMRDRREIBIGYAFVBAVDQGVHMTGTASVLLHLDKNDNPKPFTK-L 2916
QY 615 YTVRLNEDAVGTSVTVSAVDRD--AHSVITTOITSGNTRRPSITSQGGGLVSLALP 672
Db 2917 FSLVTEAEAEIGSFVIRVTSDDLIGANANASYSF--SENPKFRIEPQGN--ITVAGH 2973
QY 673 LDYTLERQYVLAVVASDGTRODTAQIYVNTDANTHAPVQSSHYTVANNEDRPAQTIV 732
Db 2974 LDRQOQOBYILIKVAASDGMARAEPTITITODONMAPEFHSFYSPFELQOSIALVG 3033
QY 733 LISATDSD--TGEMARITYFMEDSIPOPRIDADTG-----AVT 768
Db 3034 QIATDDBKQPNVSVISYLOQSPFMSIDPATSEVSKAVRKHQYVARSPEMMYALT 3093
QY 769 TOA----- 771
Db 3094 VLATDNKRPPLYSECLVINIVDAHNNPKFEQAELYALPQDAVGRQIRVHANDKOD 3153
QY 772 -----ELUYE-----DOVSYTLATIPARNGIPOKSDTT 799
Db 3154 LGTNEBMSYLMTFNLSSIPSYGRHDGMITLVKPIQVBPNTREYLVNATRGVPOQDET 3213
QY 800 YLEILVNDVNDNAPOFLRDSYQSVEDVPEFTSVLOISATDRDSGNRFRVYPOGDD 859
Db 3214 RVIVTVGEMNDPRFVNSQVIVPENEPVGTLTVGATDDDTGNGMLRHSISGANE 3273
QY 860 GGDGFIYESTSGIVRTLRIDRENAQVILRAVAVDKMPARTMEVTVTLVDVNDP 919
Db 3274 RO-DPSVDERGTGIVIOQLDYLDIQYHLNITVQDLGYHPLSSVAMLTITLTDVNDNP 3332
QY 920 VFGODEEDVFEENS----- 935
Db 3333 VFNHKEHYCIYENKPVGTVPQAHAAADKOSPKNALIHVAFPSGPDHRHFFIMQNGTI 3392
QY 936 ----- 935
Db 3393 SSAVSPEYEBRITYTLQIKAKNPSSMESYANLVHVLGNVEFPQPLQVFEHFDVSETS 3452
QY 936 -IGLAVARVATPDDEG----- 951
Db 3453 AVGTFRVAVOATDKDGEDGRVYVLLVSSNDKGRIDTNTGLIYARHLDRQNRVVL 3512
QY 952 ----- 951
Db 3513 TVNAKANYGIRGNDTDEAQVITISIQDNDEPEFIKHYYTSTISBAFVGTKVTYVKAIDK 3572
QY 952 -----TNAQIMQVIEGNIPEVFOGLDIPSGELTALVDLDYEDREPVVVIQATSA---PLV 1004
Db 3573 DVRTQNNQFYSIISNGMLKOSFKIDVOTGISTASRLDRREBTSTYMLVIGAITDGLPOT 3632
QY 1005 SRATVHVLJDRDNDPVL---GNFELLNNVYTNSSSPGGAIGRVPADPDI---SD 1058
Db 3633 GSATVHIELEDDVNDNGPFTPEG---LNGYISENBP--GTSIMTLASDPLPRNGG 3685
QY 1059 SLTYSFPERGELSLVILNASTG----- 1080

```

```

Db 3686 PFTYOLIGKHKHSLVDRNSGVVRSSTSPREBMTPIIEALIEVEDSGKPKQSGHLLTI 3745
QY 1081 -----E 1081
Db 3746 TVLDQNDNPTTSLHIAVSLFNGDLPSNVKLDVDRNDIDIVGDYRCRLQKNAPSOLO 3805
QY 1082 LKLSRALD-----NNRPLEAIMSVLVSDBGVH--SVTAOCALRVTLITDEMLTHSITLRLED 1135
Db 3806 LAIPRACDLITTSHTTPIASVPSYTGNDGHHGVSSKVSVAFQSPNNETLANSVISIVRN 3865
QY 1136 MSPERFLSPILGFIQAAVATLATPPDVVVFVNOJRTDAPGGHILNVSL-----S 1186
Db 3866 MTAYHFLANNHYRPILEMIKSRMSN--EDREVILYSI---LEGSGNSTVLOLMAVRLAKTS 3921
QY 1187 VGOBPGGGRPPPLPSDLOERLYLNSLLTALSACVLPFDNNICLRCECENTMRCSV 1246
Db 3922 YQDP-----KTLIERLRKRSASBELQKEVIVGYEPCSEPDVCENGVCSAT 3969
QY 1247 LRPDSAPFLASSS---VLFPRPIHVGILRCRCPPTGPD----- 1283
Db 3970 MRLLDHSFVIOBSPALVLSGP--RVHNDYSCQCTSGSGBQCSRRDPCLPNCHSOVOC 4028
QY 1284 -----YCEB--VDLCYSRPGCPHGR--RSBEG--GYTCLCRDGYTGEH 1322
Db 4029 RLUGSPQCMCPANBDKHEKERSDVCSKPCRNNGSCORSPDSSYFCLCRPGFRNQ 4088
QY 1323 CEVSARSGRCTPGVCKNGGTCVNLVGVGFCDCPSDPEKPYQVYTRSPHASFITFRG 1382
Db 4089 CE--SVSDSCRPNPCLHGGCLVS--LKPQKCNCTPGYGG--HCBRSYGGPQPLSYMTFPA 4144
QY 1383 LRORFHTLALSPATKRDGLLYNGRFNE--KHDFVLALEVOIVOLTFASAGSTTVS 1440
Db 4145 LDVTN--DISVATTKPNSLLIYNGMQSGSDPLATELVNGRAY--FSSGARPAIS 4201
QY 1441 PVPFG--GVSDQWHTYQLKTKNPKLLGQGLFQGPSEBQYAAVTVDCDTGVALRFGSVL 1499
Db 4202 TVIAGRLADGWHKV-----TATRNG-----RWSLSVAKCADSGVCTECLP 4245
QY 1500 GNYSCAAQ-----GTQGSKKSLDLTGPLLIGV-----PDLPSFVYRNRQFVGMRLQ 1550
Db 4246 GDSCTADREVGPVGTINFNQ-----PLMTIGLSADPLTERPGOVHSDDLVGLHSHV 4299
QY 1551 VDSRHIDMADFIANNQTVPGCPAKKNVCD---SNTCHN--GTCVNAQMDAFSCCEC----- 1600
Db 4300 IGRALNLSLPLOQKILAGC--NRQACQPALAERCGGPAQCIDRMSSSLCQCGHILQ 4357
QY 1601 -----PLGFGKSCAQ---EWANPOHFLG---SSLVAV-----HG 1629
Db 4358 SPDQSDLSPEITLIGBAPFEPRISEIYRMQLLDNLNYSKSAVLDNQWRERRAVNFS 4417
QY 1630 LSLPISQWYLSIMPRTRQADGVLOAIRGNSITLQIREGHV-----LSVBGTGLQ 1683
Db 4418 ASQIYAPKOLSLFTFYQOQGITLYAATNQMT--SLSEBGLVYYSKHOLITNTVOE 4476
QY 1684 ASLSLEPGRANDGWHHAQLALGASGPGHAILSFYQOQRA-----EG-NLGPRLH-- 1735
Db 4477 TSTL-----NDGKHNVS-----FESNSLRILVGRQVGDLDLA 4513
QY 1736 GLH-----LSNITVGGIPGAGVARGRCIQGVSDTPEGVN---SLDPS---HGB 1783
Db 4514 GYHDFLDPYLTILNVG-----EAFVGCIANTVNNELQPLNGSGSIFPEVAYHOK 4564
QY 1784 SINVEQGC 1791
Db 4565 ---IESGC 4569

```

RESULT 11  
 PATH HUMAN  
 ID PATH HUMAN  
 AC 014517;  
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cadherin-related tumor suppressor homolog precursor (Fat protein  
 DE homolog).  
 GN FAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymphocytes;  
 RX MEDLINE=96163873; PubMed=8586420;  
 RA Dunne J., Hanby A.M., Poulson R., Jones T.A., Sheer D., Chin W.G.,  
 RA Da S.M., Zhao Q., Beverley P.C.L., Owen M.J.;  
 RT "Molecular cloning and tissue expression of FAT, the human homologue  
 RT of the Drosophila fat gene that is located on chromosome 4q34-q35 and  
 RT encodes a putative adhesion molecule.";  
 RL Genomics 30:207-223(1995).  
 CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY EPITHELIAL AND SOME  
 CC ENDOTHELIAL AND SMOOTH MUSCLE CELLS.  
 CC -1- SIMILARITY: Contains 5 EGF-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin G-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X87241; CAA60685.1; -.  
 DR HSSP; P01132; IEGF.  
 DR Genew; HGNC:3595; FAT.  
 DR MIM; 600976; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0008181; F:tumor suppressor; TAS.  
 DR GO; GO:0007155; P:cell adhesion; TAS.  
 DR GO; GO:0007267; P:cell-cell signaling; TAS.  
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF 2.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00028; cadherin\_33.  
 DR Pfam; PF00008; EGF\_5.  
 DR Pfam; PF00054; laminin\_G\_1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 31.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00282; LamG\_1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 16.  
 DR PROSITE; PS00268; CADHERIN\_2; 33.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 1.  
 KW Cell adhesion; Signal; Glycoprotein; Transmembrane; Calcium-binding;  
 KW Repeat; EGF-like domain.  
 FT SIGNAL.  
 FT CHAIN.  
 FT 1 21 POTENTIAL.  
 FT 22 4590 CADHERIN-RELATED TUMOR SUPPRESSOR  
 FT 22 4183 HOMOLOG.  
 FT 4184 4204 EXTRACELLULAR (POTENTIAL).  
 FT 4205 4590 CYTOPLASMIC (POTENTIAL).  
 FT 22 149 CADHERIN 1.

FT DOMAIN 150 256 CADHERIN 2.  
 FT DOMAIN 257 361 CADHERIN 3.  
 FT DOMAIN 362 463 CADHERIN 4.  
 FT DOMAIN 464 569 CADHERIN 5.  
 FT DOMAIN 570 716 CADHERIN 6.  
 FT DOMAIN 717 822 CADHERIN 7.  
 FT DOMAIN 823 927 CADHERIN 8.  
 FT DOMAIN 928 1034 CADHERIN 9.  
 FT DOMAIN 1035 1138 CADHERIN 10.  
 FT DOMAIN 1139 1245 CADHERIN 11.  
 FT DOMAIN 1246 1345 CADHERIN 12.  
 FT DOMAIN 1346 1456 CADHERIN 13.  
 FT DOMAIN 1457 1562 CADHERIN 14.  
 FT DOMAIN 1563 1670 CADHERIN 15.  
 FT DOMAIN 1671 1769 CADHERIN 16.  
 FT DOMAIN 1770 1882 CADHERIN 17.  
 FT DOMAIN 1883 1982 CADHERIN 18.  
 FT DOMAIN 1983 2084 CADHERIN 19.  
 FT DOMAIN 2085 2185 CADHERIN 20.  
 FT DOMAIN 2186 2286 CADHERIN 21.  
 FT DOMAIN 2287 2393 CADHERIN 22.  
 FT DOMAIN 2394 2495 CADHERIN 23.  
 FT DOMAIN 2496 2599 CADHERIN 24.  
 FT DOMAIN 2600 2705 CADHERIN 25.  
 FT DOMAIN 2706 2811 CADHERIN 26.  
 FT DOMAIN 2812 2920 CADHERIN 27.  
 FT DOMAIN 2921 3024 CADHERIN 28.  
 FT DOMAIN 3025 3127 CADHERIN 29.  
 FT DOMAIN 3128 3232 CADHERIN 30.  
 FT DOMAIN 3233 3337 CADHERIN 31.  
 FT DOMAIN 3338 3442 CADHERIN 32.  
 FT DOMAIN 3443 3546 CADHERIN 33.  
 FT DOMAIN 3547 3649 CADHERIN 34.  
 FT DOMAIN 3790 3827 EGF-LIKE 1.  
 FT DOMAIN 3831 4011 LAMININ G-LIKE.  
 FT DOMAIN 4013 4051 EGF-LIKE 2.  
 FT DOMAIN 4052 4089 EGF-LIKE 3.  
 FT DOMAIN 4090 4126 EGF-LIKE 4.  
 FT DOMAIN 4127 4163 EGF-LIKE 5.  
 FT CARBOHYD 333 333 (POTENTIAL).  
 FT CARBOHYD 660 660 (POTENTIAL).  
 FT CARBOHYD 740 740 (POTENTIAL).  
 FT CARBOHYD 791 791 (POTENTIAL).  
 FT CARBOHYD 998 998 (POTENTIAL).  
 FT CARBOHYD 1426 1426 (POTENTIAL).  
 FT CARBOHYD 1551 1551 (POTENTIAL).  
 FT CARBOHYD 1751 1751 (POTENTIAL).  
 FT CARBOHYD 1867 1867 (POTENTIAL).  
 FT CARBOHYD 1905 1905 (POTENTIAL).  
 FT CARBOHYD 1943 1943 (POTENTIAL).  
 FT CARBOHYD 1994 1994 (POTENTIAL).  
 FT CARBOHYD 2328 2328 (POTENTIAL).  
 FT CARBOHYD 2467 2467 (POTENTIAL).  
 FT CARBOHYD 3326 3326 (POTENTIAL).  
 FT CARBOHYD 3424 3424 (POTENTIAL).  
 FT CARBOHYD 3446 3446 (POTENTIAL).  
 FT CARBOHYD 3615 3615 (POTENTIAL).  
 FT CARBOHYD 3642 3642 (POTENTIAL).  
 FT CARBOHYD 3718 3718 (POTENTIAL).  
 FT CARBOHYD 4154 4154 (POTENTIAL).  
 SQ SEQUENCE 4590 AA; 506273 MW; 04483CDD00E0A7 CRC64;

Query Match 9.8%; Score 1516.5; DB 1; Length 4590;  
 Best Local Similarity 24.7%; Pred. No. 2.5e-64;  
 Matches 500; Conservative 267; Mismatches 671; Indels 589; Gaps 59;

Oy 124 SPQKLTLPREHPCLKAPRLRCOSCKLAQAPGL-RAGEBSPBSLGRKRRNVYTAQFO 182  
 Db 2445 SANGITLSTNTHRAKLMP---FYSLNTSVSDGVFRSSTGVHTVYIGG---NLH-SPAF 2496  
 Oy 183 PPSYQATVPENPQAGTPVALRAIDPDGBAGRLLEYTMALFDBRSRQPSFLDPVTGAVT 242

Db 2497 QNEVEVLAENAPDLHVLWEVKTDDSGIYGVTHI--VNDPANDRYINE--RGQIF 2552  
 QY 243 TAEELRETKSTYVF--RYTAQDHGMPRBSALATLTLVTDNDHDPVFEQCKESLRE 300  
 Db 2553 TLEKLRFPFAEAVISRLMAKQAG--GKVAFCVAVIITDDNDNAPOFPAATYEVANIGS 2610  
 QY 301 NLEVEVETLVRAITDGDAPNPANILYRLLEGSGSGSPSEVEIDRSGVIRTRGPVDREY 360  
 Db 2611 SAAKGSVYV--SASDDEGSGNADITY-AIRADESEVEMENIKTSGVITTKESLIGLEN 2668  
 QY 361 ESYQVLTAEASDQDRDGPRTSTAAPLVSEDDNDNAPOSEKRYVQVREDYTPGAPVLR 420  
 Db 2669 EFTTFVRADNOSP--SKESVVLVYVKIIPPEMOLPKFSEPPYITVSDVAVGIEIDL 2726  
 QY 421 VTASDRDKGSNAVHYISMSGNA-----RGQFYLDAGTALDVVSPUDYETTK--EYTLR 473  
 Db 2727 IRRE-----HSGVVLXSLVKGNTPESNRDSFVIDQSGRLKLEKSLDHTTKYQPSIL 2781  
 QY 474 VRAQDGRPPLSVNSGLVTVQVLDINDNAPIVSTPPQATVLESVPLGVILVHQAIDAD 533  
 Db 2782 ARCTQDHEWVASVD--VSIQVADANDNSPVFESSPEAFIVENLPGSGRVIQIRASDAD 2839  
 QY 534 AGDNARLEAYRL--AYGHDPPPTINGNCWISVALEDRERVDYPSFGVRAADHGT-P-A 589  
 Db 2840 SGNGVMTSLDOSQSEVIESFAINMETGMITTLKELEKRDYQIKVVASDHEKIQ 2899  
 QY 590 LITASASVTVLVDNNDNPTFTQPEYTVRLNEDAVGTSVTVSAVDRASHV---ITYQ 646  
 Db 2900 LSTALVDVTVVDVNSPREFTAIRKGYSEDDPGGVYAIISTDDASEIRNRYTF 2959  
 QY 647 ITSGNTRNRPSTISQSGGLVSLALPLDYKLEROYLAVATSDGTRODTAQIVVNTDAN 706  
 Db 2960 ITGGDELCPFAVETIQNEMKVYVKEKLEDEKRDNYLITTAOTGTSSKAIYEVKULAN 3019  
 QY 707 THRPVQSSHHTYVNWEDRPAGTTVVLISATDEDTENARITY--FHEDESIPOIRIADNG 765  
 Db 3020 DMSVPEKTLSTPTIEDVLPGLIMQISATDADIRSNARITYTLGSGAEKRPALPDG 3079  
 QY 766 AVTTQALEYEDQVSYTLAITARDNG--IPQSDTYLELIVVDVNDNAPQFLRDSYQGS 823  
 Db 3080 ELKTSFPLRREBAVHLVLRATDGGRRPQASIVTLE---DVANDNPEFSADYAT 3135  
 QY 824 VVEDVPPFVSLOISATDRDSCINGRVTTPQGGDDGDDGFVSTSGIVRTLRIDREN 883  
 Db 3136 VVENETPRTLTFRVQATDADAGLRKILYSLI--DBADQOFSINELSGIIOLEKPLDREL 3193  
 QY 884 VAGYVLRAYAVDKMPARTPM--EYTVTVLVDVNDNPPVFEQDSFDVFEVENSPIGLAVAR 942  
 Db 3194 QAVYTLSLKAVDDGLFRRLTANGTVIVSVLDINDNPPVEYRKYGATVSEBDILVGTVALQ 3253  
 QY 943 VTATDEEGTNAQIMTQIVEGNIPFVQDLIFSGELTALVDLDYEDRPEYVLVIQAT--- 999  
 Db 3254 VYAAASDIENAMETYSIISGNHKGFSIDSKGAFIIEINDYBESHYYLVRAITDGG 3313  
 QY 1000 ----- 999  
 Db 3314 TPEISDAVAVNVATDINDNTPVFSQDTYTVTISDAVLEQSVITWADADGSPNSHIH 3373  
 QY 1000 -----SAPVSRATYHVRLL 1014  
 Db 3374 YSIIIDNGSSFTIDPVRGKVTKLLDBRTISGYTLVQASDNGSPPVNNTTVINIDS 3433  
 QY 1015 DRNDNPVL--GNFELFNN----- 1032  
 Db 3434 DVANDNAPVSRKGYSVIIDENKPVGSVLQVVTBDSHNGPPFFITVGNDEKAFEV 3493  
 QY 1033 -----Y 1033  
 Db 3494 NPOGVLLTSSAIRKKEKHVLLQVKAADNGKPOLSSLYIIRVIEESTYPPALILEIF 3553  
 QY 1034 VTNRSASPFGAIGRAVPAHDPOISDSLTVSPERGNELSLVILNASTGELKLRALDNNRP 1093  
 Db 3554 ITSSGEEYGGVIGKIHATDQDVYTLTYSLDPQMD--NLPVSSYTGKLIANKKLDIG-- 3610

QY 1094 LEAINSVLSDGSHVSTQACALRVTLITDBMLTHSITLRLIEMSPERPLSLGLFIQAV 1153  
 Db 3611 -OYLAVSTYDCKFTTADITVHIROVTOEMNLITAIAPANLITBEFVGDMRNFORL 3669  
 QY 1154 AATLATPPDHVVVFNVDTPDAPGGHILNVSLVSGPPGPGGPPPLPSEDLQERLYLNR 1213  
 Db 3670 RNILGVRRDDIQLVISIQ--SSBPHH-LDVLLFVEKP-----GSAQISTKQLHKNSSV 3721  
 QY 1214 SULTIASAORVLPFPDNLCLREPCENYKCVSLAPDSAPFIASSVL--FRPLHPVG 1271  
 Db 3722 TDIEHIGVRLINVPQKLCAGIDCP--KRCDEKAVSDSVMSTHSTARLSFVTPRHRAA 3780  
 QY 1272 LRC-----RCPPFTGADVCETEVDCYSRPGCHRCR--EGGYTCLCRDGYGHEC 1324  
 Db 3781 V-CLCKEGRCPVNHG-----CEDDPCRGSECTVSDPHEKHTCYCSG----- 3823  
 QY 1325 VSARSGRCTPGVCKXNGGTCVNLVGFKDCDPSGDPFEKPYCOVTRSPRASHFITER--G 1382  
 Db 3824 ---RFGQC-PG-----SSGMTLTGNSYVXKRLTE 3848  
 QY 1383 LRGPHFTLASFATKEDGLLYNGRFRNEKDFVALSVIQOVQLTSAGSITTVSBEF 1442  
 Db 3849 NENKLEMLTWMLRTYSTHAAVMY---ARGTYSILIHGRRLQYKFCDSG----- 3897  
 QY 1443 VEGVS-----DQOMHTVOLKYYNKPRLGQYGLPQGSBQKAVVTVDGCOTGVALRF 1495  
 Db 3898 -PGIYSVGSIQVNDQMANVALL----- 3919  
 QY 1496 GSVLGNYS-----CAAQSTQSGSKSLDTGLPLLGQVDPDPSFVYMRQ----- 1541  
 Db 3920 --VNGNVARLVLDQVHTASGTAPGLTKTLINDNYVFPFG-----HROQSTRHOR 3967  
 QY 1542 -----FVCGM-----RULQVDSRIIDMADPFIANNGYPGC--PAKANVCBNTGKN 1585  
 Db 3968 SPQVNGFRGCDISTYLNGQELPLNSKRSYAHIEESVDVSGCFPLATED--CASNPCCN 4026  
 QY 1586 GGTG--VNQMDAFSCBCEPLGFGSKSCAQEMANPQHPLGSSLVAMHGLSLPISQPYLILMF 1644  
 Db 4027 GGVCNPPSAGGYCKCSALYIGTIC----- 4051  
 QY 1645 RTRQADGVLLAITRGRSTITLQLREHVMLSVBGTGLQASSLRLEPGRANDGDMHNAOL 1704  
 Db 4052 -----EISVNPCSSN----- 4061  
 QY 1705 ALGASGGRGHMILSPDYQGRABGNLGRHLHLSNTVGGIIPRAGVARGFPGCLOG 1764  
 Db 4062 -----PCLYG-----GTCVVDNGAFVCCQCKGLYTG 4086  
 QY 1765 VRVSDTPREGVNSLDPSSHGESINVEQCSLPDPCDNPSPANSYCSNDMDSYSCSDPGY 1824  
 Db 4087 -----QRCOLSPYCXDEPRCKNGTCDSDLDGAVCCQDSGFR 4122  
 QY 1825 GDNQ---TNVCDLNPCHQSVCTRKPSAPRG-YTCECPNYLPGYCE 1867  
 Db 4123 GERCOSDIDECSGNPLGALCEN---THGSYHNCCHERYGRHCE 4165  
 RESULT 12  
 FAT2\_DROME STANDARD; PRT; 4705 AA.  
 AC Q9VW71; Q95S51;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Puctive fat-like cadherin-related tumor suppressor homolog  
 DE precursor.  
 GN FAT2 OR CG7749.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 OK NCBI\_taxid=7227;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoeklin R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame C.R., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abil U.F., Agbayani A., An H.-J., Andrews-Plannkoc C., Baldwin D.,
RA Ballew R.W., Baen A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Butkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Decher A., Deng Z., Nays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Idegawa C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milihina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrakos R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP REVISIONS.
RC STRAIN=Berkley;
RX MEDLINE=22426069; PubMed=12537572;
RA Miara S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield B.U., Bayraktaroglu L., Bertman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [3]
RP SEQUENCE OF 3837-4705 FROM N.A.
RC STRAIN=Berkley; TISSUE=Ovary;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Chame M.,
RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- SIMILARITY: Contains 34 cadherin domains.
CC -1- SIMILARITY: Contains 5 EGF-like domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003515; AAF9078.2; -
DR EMBL; AY060955; AAL28503.1; ALT INIT.
DR EMBL; AY118666; AAM50035.1; ALT_INIT.
DR HSSP; P15116; INCI.
DR FLYBase; FBgn0036930; fat2.
DR GO; GO:0005887; C:integral to plasma membrane; ISS.
DR GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; ISS.
DR GO; GO:0016339; P:calcium-dependent cell-cell adhesion; ISS.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00028; cadherin; 31.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00054; laminin G; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 34.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00282; LamG; 1.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_2; 34.
DR PROSITE; PS00268; CADHERIN_2; 34.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS00025; LAM G DOMAIN; 1.
KW Hypothetical protein; Cell adhesion; Signal; Glycoprotein;
KW Transmembrane; Calcium; Calcium-binding; Repeat; EGF-like domain.
FT CHAIN 1 35
FT 36 4705
FT 1 35
FT 36 1647
FT 1648 1668
FT 1669 4705
FT 60 180
FT 181 288
FT 285 397
FT 398 504
FT 505 610
FT 611 713
FT 770 874
FT 875 977
FT 978 1085
FT 1086 1195
FT 1191 1296
FT 1297 1402
FT 1405 1503
FT 1504 1609
FT 1610 1714
FT 1715 1812
FT 1813 1929
FT 1929 2030
FT 2031 2137
FT 2138 2238
FT 2239 2338
FT 2339 2465
FT 2466 2567
FT 2568 2670
FT 2671 2779
FT 2780 2876
FT 2877 2983
FT 2984 3088
FT 3084 3185
FT 3186 3289
FT 3290 3394
FT 3395 3499
FT DOMAIN 36 1647
FT TRANSSEM 1648 1668
FT DOMAIN 1669 4705
FT DOMAIN 60 180
FT DOMAIN 181 288
FT DOMAIN 285 397
FT DOMAIN 398 504
FT DOMAIN 505 610
FT DOMAIN 611 713
FT DOMAIN 770 874
FT DOMAIN 875 977
FT DOMAIN 978 1085
FT DOMAIN 1086 1195
FT DOMAIN 1191 1296
FT DOMAIN 1297 1402
FT DOMAIN 1405 1503
FT DOMAIN 1504 1609
FT DOMAIN 1610 1714
FT DOMAIN 1715 1812
FT DOMAIN 1813 1929
FT DOMAIN 1929 2030
FT DOMAIN 2031 2137
FT DOMAIN 2138 2238
FT DOMAIN 2239 2338
FT DOMAIN 2339 2465
FT DOMAIN 2466 2567
FT DOMAIN 2568 2670
FT DOMAIN 2671 2779
FT DOMAIN 2780 2876
FT DOMAIN 2877 2983
FT DOMAIN 2984 3088
FT DOMAIN 3084 3185
FT DOMAIN 3186 3289
FT DOMAIN 3290 3394
FT DOMAIN 3395 3499
FT PUTATIVE FAT-LIKE CADHERIN-RELATED TUMOR
FT SUPPRESSOR HOMOLOG.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT CADHERIN 1.
FT CADHERIN 2.
FT CADHERIN 3.
FT CADHERIN 4.
FT CADHERIN 5.
FT CADHERIN 6.
FT CADHERIN 7.
FT CADHERIN 8.
FT CADHERIN 9.
FT CADHERIN 10.
FT CADHERIN 11.
FT CADHERIN 12.
FT CADHERIN 13.
FT CADHERIN 14.
FT CADHERIN 15.
FT CADHERIN 16.
FT CADHERIN 17.
FT CADHERIN 18.
FT CADHERIN 19.
FT CADHERIN 20.
FT CADHERIN 21.
FT CADHERIN 22.
FT CADHERIN 23.
FT CADHERIN 24.
FT CADHERIN 25.
FT CADHERIN 26.
FT CADHERIN 27.
FT CADHERIN 28.
FT CADHERIN 29.
FT CADHERIN 30.
FT CADHERIN 31.
FT CADHERIN 32.

```

```

FT DOMAIN 3500 3604 CADHERIN 33.
FT DOMAIN 3605 3712 CADHERIN 34.
FT DOMAIN 3819 3879 EGF-LIKE 1.
FT DOMAIN 3881 3919 EGF-LIKE 2.
FT DOMAIN 3937 4121 LAMININ G-LIKE.
FT DOMAIN 4129 4166 EGF-LIKE 3.
FT DOMAIN 4168 4205 EGF-LIKE 4.
FT DOMAIN 4243 4279 EGF-LIKE 5.
FT DISULFID 3823 3835 POTENTIAL.
FT DISULFID 3830 3867 POTENTIAL.
FT DISULFID 3869 3878 POTENTIAL.
FT DISULFID 3885 3896 POTENTIAL.
FT DISULFID 3890 3907 POTENTIAL.
FT DISULFID 3909 3918 POTENTIAL.
FT DISULFID 4133 4144 POTENTIAL.
FT DISULFID 4138 4154 POTENTIAL.
FT DISULFID 4156 4165 POTENTIAL.
FT DISULFID 4172 4183 POTENTIAL.
FT DISULFID 4177 4193 POTENTIAL.
FT DISULFID 4195 4204 POTENTIAL.
FT DISULFID 4247 4258 POTENTIAL.
FT DISULFID 4252 4267 POTENTIAL.
FT DISULFID 4269 4278 POTENTIAL.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 923 923 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1198 1198 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1312 1312 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1473 1473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1511 1511 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 3962 3962 G -> E (IN REF. 3; AAL28503).
SQ SEQUENCE 4705 AA; 524564 MW; 6D387A489D2C3DR CRC64;

```

Query Match 9.5%; Score 1477.5; DB 1; Length 4705;  
 Best local similarity 22.8%; Pred. No. 1.9e-62;  
 Matches 462; Conservative 282; Mismatches 693; Indels 593; Gaps 53;

```

QY 178 APOPPSPQATYPRNOPAGTPVSLRAIDPRDGEAGRLTYMDALPDRSNQFSLDPV 237
DB 2564 SPFDOSTYEAQVPEMLHGHNIITVKAISDGDGTANYLY---EIVSEMKKIFLIDQT 2620
QY 238 TGAUTTAELDRRTKSTHVRVTAADHGMFRASALATLTLYTDNDHPVEEQGYKKS 297
DB 2621 TGVITKVTFRKAKDEIVLVLVKVSOGG--GKFGPASKVTVVDNDVPEFLKEIKKV 2678
QY 298 LRENLEVGVEVLTVRAIDGDAPNPANILYRLLEGSGG--SPSEVFEIDPRSGVIRTRGPV 356
DB 2679 VSTTVANQGITLVKAKDDIDVNGSVHPQIVOKSKMDKAVKQVIEINERTGDIVFRSKAE 2738
QY 357 REVEVSQULTVKSODGR-----DPG--- 377
DB 2739 SYGVNSYQFVRASDGEPPQFHSSEVPSIEIETDANIPFKFSVYLKIIESTPFGTVL 2798
QY 378 -----PRSTTAVF--- 386
DB 2799 TKLHMIGNITFKSIAADQHFMIISGELLLOQTTLDREQESHNLIIVAEISTVVFVFA 2858
QY 387 -----LSVEDNDNAPQFSEKRYVVOVREDVTPGAPVLRTVADSRRKGSNAVHVSIMG 441
DB 2859 YADVLDIVDRDNDNYPKFDWTFYSASVAENSEKVISLVKVSATDADTGPNGDIRLYLESD 2918
QY 442 --NARQGYLDAQTGLDVSPIDYETTKGYTLRVAQDQGRPLSNVSGLVTVQVLDIN 499
DB 2919 TENIQIFIDIDIVSGITLTSLDREVOSEYKFKVLAADNGHK--DPAKPYVIKIKVDVN 2977
QY 500 DNAPIVSTFPPQA--TVLESVPLGLVYLVAQIDADAGDNARLEFRLAGVGHDPPTLNNG 558
DB 2978 DNAPVF--KLPISGLSVFENALPQTIVLINTLLIDPDI--EKQEMDFIVSGDKQAQFOIGK-- 3034

```

```

QY 559 TGMISVAELDEBEVDVFSFGVEARDHGTALTASASVTVLDVNDNPTFOPEYTVR 618
DB 3035 SEGELFAKLDEBQMLFTYLSIIAAD---GKTPARANVIEDVNDINDNTPYCKAPRYHS 3091
QY 619 LNEDAAVGTSVVTVASANDDAHSVITTYQTSGRTRRFSITSQSGGLVSLALPDYKLE 678
DB 3092 TRESISIGTLVEVKAKIDFDFOSKLRFLY--SGKADDFSIGKES--GILKVASALDREFT 3148
QY 679 RQVLAVTSAG---TRODTAQIVNVTNATNHRPFGSSHVTYVNNERPACTVVLIS 735
DB 3149 PKYKVAHVODGDQDFQBCFSEIITVNDINDNMPFISMAQYRVSEDAQIANTLTAKVH 3208
QY 736 ATDEPDGENARITYFV----- 751
DB 3209 AADKQDGVNRQIKTISLGMENHDYFKISKSTGILRLHKSIDRETISLFLNLTVAEDCGVKR 3268
QY 752 -----EDSIPO----- 757
DB 3269 LHSIATVAVNIIIDINDNPEPFEMRQYSCILRNATHTGEVCKVYATSIDIGNADIHFI 3328
QY 758 -----PRIDADGAVTTOAEIDYEDQVSTLTAITADNGIPQKSDTTYLEIVNDVN 809
DB 3329 MEGNEQKFKMDSTGDLVLAATLDYEMSKFYLTIOAIDGTPPLSNNAVYNISILDIN 3388
QY 810 DNAPQFLRDSYQSYVEDVPFTSVLQISATDRDGLNGRVYTFQGGDGDGFVSEST 869
DB 3389 DMSPTFLQVLVANNEDIVFVSKLIDVATKDESDVNGLVYNNIRGN--IGQSIDXK 3447
QY 870 SGIVTTLRLRENNVAQVYLAAYAVDKMPARTMEVTVTVLDVNDNPFVEODEFDVF 929
DB 3448 NCTISGRPLDRETSHTYLEIOACDQGPQGCNSVPININILTDNDNAPIFSSNSYSV 3507
QY 930 VENSPIGLAVARVATPDDEGTN--AQIMQVBEINPEVPODIFSGELTALVDLYD 988
DB 3508 IQENMLIGVLTFTFKISDADTEFTNTPTFTDRSNBEGFLFRL--QDGLRTASRNNHL 3566
QY 989 REPEYVL---VIGQASPLVSRATVAVRLDRNDNPVLGNFELFNNTYNNRSSPFGA 1045
DB 3567 QDEFVIOVRVPNGRPPLYSADMVYVKIIBEQYRPITVPLEVTINSF---EDDFSGAF 3622
QY 1046 IGRVPAHDPDSDSLTYSEFEGNE--LSVLILANS--TGBIKLSRALDNNRPLEAINSV 1100
DB 3623 ICKVNASDQDKYDELNFSLVSGPDMDYOSKLFNINNTYTKAYIASNLDIG--YKLVN 3679
QY 1101 LVSQVHSTVACALRVITIDEMLTHTSLTLEMSDERFLSPILGFIQAVATIANP 1160
DB 3680 SVSDGKTFVFSIVKINVELVNDMLKESVVIKFRPISASEFLSHRKTFRKIRNIMCR 3739
QY 1161 PDHVVVFNVQRTDAPAGHI-----LNVSLVGGPPGPGGPPFLPSEDLQER 1208
DB 3740 QKDVLITLQSDYQKASQAHAVGNRRASIDSLANVVPVVRKQIILPDDBFTSTBEINRT 3799
QY 1209 LYIARSL--TALSAQVLP---FDNITCLAREPCENYMRVCS--VLRPSSAPFLAS 1258
DB 3800 LIDKNGEIEENENLVVEBVLVSTCOGNKNDCHGRCOKIQLIKNNVTTFTFDVISAAP 3859
QY 1259 SSVLFRPHVPGGLRCRCPGPTGDCYCEBVLDLCVSRGCPGRGRSREG--GYCLORD 1316
DB 3860 SYI-----PVN--TCVCRPGDGRHCKEYVACSTDPCSFRICMPBGSAIGYCVCPK 3911
QY 1317 GYTGHECVSASRGTCPGVCKNGGTGVNLLVGFKCCPSPGDFEKPVCQVTRTSFPFAS 1376
DB 3912 GFSGTyce--RKSXCSNBSOMG-----LPTAVSFGGKS 3944
QY 1377 FTFRGLRORFFTY-----ALSPATKRDGLLYNKRFRNEKHDPAVALVYQEOVLT 1430
DB 3945 YAHYKINKKAKFTLENGFSYSIQIRTVQQTGLLY--ASGKVDYNNILEITNGAVQYRF 4001
QY 1431 --SAGESITVSPFVPGVSDQOMHTVOLKYNNKLLGOTGLPQSPSQKAAVTVVGDG 1488
DB 4002 DUGSBEGVISYST---NISDGMHQISLER----- 4029

```

QY 1489 TGVALRFSVLTGANTSCAAQGTGGSKSLDTLTPILLGCVDPLE---SPVPRNQFVG 1544  
 Db 4030 ---SINSKXVWMDNGVHSGAPGNGILNIOQNDIFVGAERVPHPSITIGEDIORFGFI 4086  
 QY 1545 GMRNLOVSRHIDMADFIANNGTVPACPAKNNV-----CDSENTCHNGTGC 1589  
 Db 4087 CMANIKIKESIPL--YISGOSTTAAKRFTNVEKCPDSPNVLVLGICSSOPCANSIC 4144  
 QY 1590 VN-QMDAFSCCEPLFGGKSCAQMENVNPHFLSGSLVAMHGLSLPISQPMYLSIMFTRQ 1648  
 Db 4145 KELTDVDECAACQPRYSKGKCEIDL----- 4159  
 QY 1649 ADGVLQAITGRGRITITQLRERGVMSVEGTGLQASLRLEPRGRANDGWHHAQALGA 1708  
 Db 4170 ----- 4169  
 QY 1709 SGGPCHALISFDYGGQRAEGLNLPRLHGLHLNITVGI PGPAGGVANGPFGCIGYRV 1768  
 Db 4170 ----- 4169  
 QY 1769 DTPEGVSLDPSHGESINVEQCSLPDPCDNPCCPANSYCS-NMWSYSCSDRQYVDN 1827  
 Db 4170 -----DCCSSGSPCLFGRCQDYHGNPNYSCTCPHLSGKR 4203  
 QY 1828 CT--NVCIDNCEHOSVCTRKPASAPKGTCECPNYLGPYCTETRID---QPCRGMWGH 1861  
 Db 4204 CEYKGFCTPNPCCKNGKIGEBSDGISH---CMC-RGYTGTPTETIDVDECNQPCNG- 4255  
 QY 1882 PTGFCPCNDVSKGFPDPCNKTSGECHKENHYRPPGSPYTCCLCPCPYGS 1931  
 Db 4256 ATC-----INERGSFRC-ICRSTYLTGA 4276

## RESULT 13

CADN\_DROME STANDARD; PRT; 3097 AA.  
 AC 015943: 09VJ37; Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neural-cadherin precursor (Cadherin-N protein) (DN-cadherin).  
 GN CADN OR CG7100.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM D).  
 RC TISSUE=Embryo, and Head;  
 RX MEDLINE=97388431; PubMed=9247265;  
 RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;  
 RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion  
 RT receptor, in the Drosophila embryonic CNS.";  
 RL Neuron 19:77-89(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baxa A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan D.A., Bouck J., Brokstein P., Brotler P.,  
 RA Buttle K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fobler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Gloder A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris K.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin M., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jatalil M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Melnikov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Berkley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Belletcourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Ruben S., Schroeder W.M., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP INTERACTION WITH ARM.  
 RX MEDLINE=98298928; PubMed=9635189;  
 RA Loureiro J., Pelifer M.;  
 RT "Roles of Armadillo, a Drosophila catenin, during central nervous  
 RT system development.";  
 RL Curr. Biol. 8:622-632(1998).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL  
 CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL  
 CC INFORMATION.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=D;  
 CC IsoId=O15943-1; Sequence=Displayed;  
 CC Name=A;  
 CC IsoId=O15943-2; Sequence=VSP\_000667, VSP\_000668;  
 CC Name=B;  
 CC IsoId=O15943-3; Sequence=VSP\_000668;  
 CC Name=C;  
 CC IsoId=O15943-4; Sequence=VSP\_000667, VSP\_000668, VSP\_000669;  
 CC Name=E;  
 CC IsoId=O15943-5; Sequence=VSP\_000667;  
 CC Name=F;  
 CC IsoId=O15943-6; Sequence=VSP\_000669;  
 CC Name=G;  
 CC IsoId=O15943-7; Sequence=VSP\_000667, VSP\_000669;  
 CC Name=H;  
 CC IsoId=O15943-8; Sequence=VSP\_000668, VSP\_000669;  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN





```

QY 772 ----- 771
Db 1733 KDINDMPOQTKDEWFEVDETDGTALPEMPLITVYVHDEDETNKFQYKIDNSGYADK 1792
QY 772 ----- 771
Db 1793 FTWVNNNDGTSGLKIVQPLDYEDDQLOSNQFRFRQVNDKGBENDNDKXVAVSWVAVFLR 1852
QY 807 DVNDNAPQFLRDSYQGSVYEDVPPTSTVLSQISATDRDSGLNGRVYTFQGGDGDGDFIV 866
Db 1853 DINDKHPFERANVSVFEDOTKVTLEKPKATTPDQGSKSYSYIDRSSDRQRAI 1912
QY 867 ESTSGIVTFLRLDRBNVAQVYLAAYVADKMPARTMEVTVTVLVNDNPNVPEQDEF 926
Db 1913 -NONSATVIOBSLDEVPVRHQVKILAIIDGSPPTATATATLVIVQINDNAPKELKQYR 1971
QY 927 DVFVENSPIGLAVAVTATDPDEBCTNAQIMQIVESGIPRVQDIFSGE----- 977
Db 1972 PVLPEHVP--RKVEIILATDDDRSKS-----NGPPQFRLDPSADDIRASFPVE 2021
QY 978 ----- 1792
Db 2022 QDQKANDGMAVIVSLRSPREQOKETMIPVIMKHGSPAMTGTSTLVIIQVNDK 2081
QY 1022 VLGNEFLLFNNTVNRSSSPFGAIGRVPAHD--DISDLTY-----SFERGELS 1071
Db 2082 QPQSKDIFVYNY-----QGSPTPIGRVYVVDLDMDLDPDKFYMEAMEHPRFKLDEDSG 2137
QY 1072 LVILNATNGE-----LTLSTRALDNNRLEAMSVLVSDGVH--SYTACALAVITTEM 1124
Db 2138 WVTMAAGREGRYHLRFVYDRKHTQTDIPANVTYVBIPIHEAVNAGSVRLSISD 2197
QY 1125 LTHSTTLLEDMSPERFLPLGLFIOVAATLATPRHVVVFNQORDTAPGGHILNS 1184
Db 2198 FIRVWNTYQMSNRK-----MDRFRDLADLINTERENVOIFSQQLRKRP--PLTVR 2250
QY 1185 LSV--QGP--PQGGGPPPLPSEDIQERLYANRSLITLAI SAQVLPDPDNCIAREPCNY 1240
Db 2251 FSANSSPYKPRVINGIIVIMHREIEKQVGIN--ITWVGIDECI--YENQMC--BGSCTNS 2305
QY 1241 MRCVSALFDSAPRIASSSVLFRITHVGLRCRCPGPGDVCETEVDLCYSRCPBH 1300
Db 2306 LFI-----SPLFVWVNN-----KTALVGRVDTIAD----- 2332
QY 1301 GRCSRREGYTCLCRDGYTGEHCEVSASGRCTPGVCNKGCTCVNLVGFKCDPCSGDF 1360
Db 2333 --C-----TCGANFTKPBSC-----RTTP--CHNGRCVDTFRPG--HCSFPGV--Y 2372
QY 1361 EKPVCQVTRSPPAHSFTTFRGLRORFHTLALSPATERDGLLYNGRFNEKH----- 1414
Db 2373 TGPRCQOTTRSPFRGNGMAYPPLMWCDSHLSLEFITTCKPDLITVNGPIVPERDETLI 2432
QY 1415 -DFVALLEIOBOVOLTFAGBSSTTVSPFVPGVSDGMHMYOLKYKPKLLGQGLRPG 1473
Db 2433 SDFILBLEKRGVPRLLIDFGSGTLELYKTKTLDGEMHRIIDLWDTESIR----- 2484
QY 1474 PSBOKVAVVTVDC-----DTGVALARFSGVLGNVSCAOGTQGSKKSLDTLGPULL 1525
Db 2485 -----MVVDFCSARIEAEMEDGTPEPDM-----SQCARQGPFPNEYLVANVAPLOV 2532
QY 1526 GGVPLPSSFPVRM-----RQFVGMRLQVDSRHIMADIANNNGVPGCPACKN 1576
Db 2533 GGL--YRQFQDQSLFWMHYMPAKPDCIRNLVNSKLYDLAHGLSRNSVAGCPQEBE 2590
QY 1577 VC-----DENTCHNGGTGVNOMDAFCECEPLFGFGKSCAOEMANPOHFGSSLVANHGSL 1632
Db 2591 VCAQETTRACMEHNCVCSLSHACHRCRPGMTGAC--NIPITPTTFAQSVYKALSF 2649
QY 1633 PISQFWYLSLWFRTRQADGVLLQALTR--GRSTITVLQBRGHV--MLSVETGLQASSLRL 1669
Db 2650 FDRFSTQVQLRRTREBYGELFRVSDQHNREYGLIIEKHGLHFRYNNLSRTEKDLML 2709
QY 1690 EPRGRANDDMHHAQALASGSGPRAIISFDYQQRABGN-----LGRPLHGLHL 1739

```

```

Db 2710 NAIIVNDGOMHVVVNNRYGSA-----ATLELDGEGREYNETFEFVGHQMLLVQKEGVY- 2764
QY 1740 SNIVGSIIPBAGVVA-----RGF-----RGLOQVRVS-----DTPEGVNSLDPSHGE 1783
Db 2765 -----AGKAAVYGVTRTEFYVADYQKSLDIRLEBKRLPLPPANNG--TQWQ 2811
QY 1784 SI---NVEQCGSLPDPDSDNPNCPANNSYCSNDMDYSQSC-----DPGYGDNCVNCYL 1834
Db 2812 ATMARNLKQGPNSKPCSNVTCPPDFECVLDLMNVYECTCGGRIMSPDSKGMORNECLD 2871
QY 1835 NPCEHQSVCIR-KPSAPHGTCBCPPNYLAPYCE 1867
Db 2872 MPCMNGATCINLEPRL--RYRCICPDGFVGNCE 2903

RESULT 14
PAT2_HUMAN
ID PAT2_HUMAN STANDARD; PRT; 4349 AA.
AC Q9NYQ8; 075091; Q9NSR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protocadherin Fac 2 precursor (hPat2) (Multiple epidermal growth
factor-like domains 1).
GN PAT2 OR MEGF1 OR CDH8.
OS Homo sapiens (Human).
OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINB:20202599; PubMed:10716726;
RA Wu O., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
feature of protocadherin genes."
RT Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 3777-4349 FROM N.A.
RX MEDLINB:98360089; PubMed:9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
EGF-like motifs by motif-trap screening."
RT Genomics 51:27-34(1998).
RN [3]
RP SEQUENCE OF 4142-4349 FROM N.A.
RX TISSUE-Testis;
RA Pouscka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 32 cadherin domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC
CC EMBL; AF231022; AAF61928.1;
CC EMBL; AB011535; BAA32463.1;
CC EMBL; AL157443; CAB75663.1;
CC PIR; T46927; T46927.
CC Genew; HGNC:3596; PAT2.
CC MIM; 604269;
CC HSSP; P00740; IEDM.
CC GO; GO:0005509; F:calcium ion binding activity; NAS.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.

```

DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF000028; cadherin; 33.  
 DR Pfam; PF00008; EGF; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 31.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00282; Lame; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 14.  
 DR PROSITE; PS00268; CADHERIN\_2; 32.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 1.  
 DR Transmembrane; Glycoprotein; Repeat; EGF-like domain; signal.  
 KM SIGNAL 1 18  
 FT CHAIN 1 4349  
 FT TRANSMEM 19 4048  
 FT TRANSMEM 4049 4069  
 FT DOMAIN 4070 4349  
 FT DOMAIN 34 148  
 FT DOMAIN 149 256  
 FT DOMAIN 363 458  
 FT DOMAIN 459 564  
 FT DOMAIN 565 669  
 FT DOMAIN 716 820  
 FT DOMAIN 821 925  
 FT DOMAIN 926 1032  
 FT DOMAIN 1033 1137  
 FT DOMAIN 1138 1242  
 FT DOMAIN 1243 1346  
 FT DOMAIN 1350 1448  
 FT DOMAIN 1449 1555  
 FT DOMAIN 1556 1660  
 FT DOMAIN 1661 1758  
 FT DOMAIN 1759 1872  
 FT DOMAIN 1969 2070  
 FT DOMAIN 2071 2171  
 FT DOMAIN 2172 2272  
 FT DOMAIN 2273 2379  
 FT DOMAIN 2380 2481  
 FT DOMAIN 2482 2585  
 FT DOMAIN 2586 2691  
 FT DOMAIN 2692 2797  
 FT DOMAIN 2798 2906  
 FT DOMAIN 2907 3011  
 FT DOMAIN 3012 3113  
 FT DOMAIN 3114 3218  
 FT DOMAIN 3219 3321  
 FT DOMAIN 3322 3426  
 FT DOMAIN 3427 3531  
 FT DOMAIN 3532 3642  
 FT DOMAIN 3643 3773  
 FT DOMAIN 3773 3944  
 FT DOMAIN 3947 3984  
 FT DOMAIN 3986 4022  
 FT DOMAIN 3297 3302  
 FT DISULFID 3951 3962  
 FT DISULFID 3956 3972  
 FT DISULFID 3974 3983  
 FT DISULFID 3990 4001  
 FT DISULFID 3995 4010  
 FT DISULFID 4012 4021  
 FT CARBOHYD 39 39  
 FT CARBOHYD 210 210  
 FT CARBOHYD 280 280  
 FT CARBOHYD 330 330  
 FT CARBOHYD 459 459  
 FT CARBOHYD 568 568  
 FT CARBOHYD 627 627  
 FT CARBOHYD 655 655  
 FT CARBOHYD 789 789  
 FT CARBOHYD 966 966  
 FT CARBOHYD 1175 1175  
 FT CARBOHYD 1303 1303  
 FT CARBOHYD 1383 1383

FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1904 1904 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1998 1998 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2007 2007 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2183 2183 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2325 2325 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2368 2368 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2387 2387 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2470 2470 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2547 2547 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2597 2597 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3125 3125 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3276 3276 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3310 3310 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3430 3430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3471 3471 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3601 3601 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3772 3772 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3813 3813 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3840 3840 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3873 3873 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3989 3989 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 4117 4117 P -> L (IN REF. 2).  
 FT CONFLICT 4160 4160 E -> G (IN REF. 2).  
 SQ SEQUENCE 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;  
 Query Match 8.5%; Score 1320.5; DB 1; Length 4349;  
 Best Local Similarity 26.1%; Pred. No. 5,8e-55;  
 Matches 478; Conservative 251; Mismatches 724; Indels 377; Gaps 65;  
 QY 175 VNT-----APQPPSYATVPENOPACTPVASLRAPIDPDGEGRLTYTMDALPDSRS 228  
 DB 2469 IYTNANKSPFQOHLYEABLAENAVGTKYIDLAIDKDSGPYCTIYT--IINKXA 2525  
 QY 229 NQFSLDPVTGAVTAEBLDRFTKSTHV--FVTAQDHMPRRSALATLTIVTDTNDH 286  
 DB 2526 SEKFSINP-NGQATLQKIDRENSTERYALAVMARDDG-GRVAFYKILITDENDP 2582  
 QY 287 PVFEQDEYKESLRLENLEVEYEVLTVAATDGPAPNANILYRLBSGSGSP---SEVFEI 342  
 DB 2583 PEFKASEYVSIQSNVSKOSPVIQVLAADADGQADVTVSV-----NPEDLVMDVIEI 2636  
 QY 343 DRSGVIRKRGVNDREKVESYQLTVEASDQGDPPGRSTTAAVFISVEDNDNAPQFSK 402  
 DB 2637 NPVTGVVKVQKDLVGLNQTLDPFITKAQDGG--PFWNSLVVVRLLQVVPKVSILPFSBP 2694  
 QY 403 RYVVOVREDVNTGAP---VLRVTSADPRKGSNAVVHYSIMSG-----NARGQFYDAQNG 454  
 DB 2695 LITFSAPEDLPBGSEIGIKVAAQD-----PYISLVKRTTSPESKQVFSLDPDIG 2747  
 QY 455 ALDVVSPLDYETTKETYLVRADGAPPLSNVSGI---VTVQVLINDNAPIFVSTPF 510  
 DB 2748 VIKVRKPMDHSTKIQXDVMNH-----CLQNTDVLSVSNVIQVGDVNDKPVFEADPY 2802  
 QY 511 QATVLESVVLGLVLAHQALDADAGDNARLEYRL-AGVG--HDPPTINNGTGWISVA 566  
 DB 2803 KAVLTENMEVGVSVIQAVALIDDTGRDGVSRLLADPQSNVHEI-FALDSRSGWITTLQ 2861  
 QY 567 ELDRREVDVDSGVGEVRPDHG--TPAL TASASVSVTVLDVNDNPTPTPEYTRVLNDAV 625  
 DB 2862 ELDCEICQYTHHVAAYDHGQITQLSSQALVSVITTDENBNAPRASESRSSVENSSEP 2921  
 QY 626 GTSVTVSAVDND--AHSVITYQITSGTNRNFSITSGGGGLVSLALPLDYKEROY 682  
 DB 2922 GELVATLKLTDADISEQRQVTCYITBGPPLQFGISQGDWRLSSRKTLDREHTAKYL 2981  
 QY 683 LAVTSADGRQDTADIVNATDANTRHPVPOSHTYANVNERPAGTVVLAISADEDNG 742  
 DB 2982 LRVTSADGRQASVVEIIFVLVDVNDNSPQSQQLYTGKVRHEDVFGHFLIKVASATDLDTD 3041

```

QY 743 ENARTTME-DSIFRIDADTGAVTTOAEIDYEDQVSYTATATARDNGIPOKSDTYL 801
D 3042 TNAQITTSIHGPGAHFCLDHTGELTTLTALDERKQVPLVAKATDGC--GSSCADT 3099
QY 802 EILVNDVNDNAPQFLRDSYQGSVEYEDVPFTSVLQISATDRDSCGLNEVFYTPQGGDGD 861
D 3100 TLHVEDVNDNAPRFPSSHCAVAVFNDTNTKTPVAVFARPDQGANQVYSLP--DSAE 3157
QY 862 GDPFVSTSGVITLRLDRDRENAQVYLRAVADKGM-PARTMEYTVTLVDNDNPPV 920
D 3158 GHFSIDATGVILREKPLQVPRQAPLELTVASDGTPIPLSTIGTIVSVVGGEDYLPV 3217
QY 921 FEODEFVFEENSPIGLAVARV-TATDPD-EGTNAQIMYOIVEGNIPEVQOLDFSGEL 978
D 3218 FLNTHSVOVEDAPDPTGVQLATLTPGAETG--YVSGNEGRRLDARLTIL 3273
QY 979 TALVDLDEYDEPEYLVIA--TSAPLVSRATVHRLDENDNPPVLANPEILFNNTYT 1035
D 3274 YVNASLDEFTSPKTYLSTECRSKSSSLDVTVMVNTDVENHRP--QPODPYST 3328
QY 1036 N-RSSSPFGAIGRPAHDPD--ISDLSITSPERGENEISVLNASTGELSRALDNMR 1092
D 3329 RVLNMAVGDIVILTVSATDEDEGPLNSDITTSIIGNQGHFTIHPKKGELQVAKALDREQ 3388
QY 1093 PLEAIMSVLSDG--VHSYTAOCALRVITITBEMLTHTSITLRLDMSPERFSLPILGL 1148
D 3389 ASSYSIKARATDSGQPLHEDT-DIAIYADVND-----NPRFQOLNTST 3433
QY 1149 FIQAVATLATPDPHVVFNQRTDADGAILNVLNLSVGPPGCGPP-- 1199
D 3434 TVQ-----ENSPGSKVLQILIS--DPDSEPNPYSFRITKGN 3471
QY 1200 -----LSEDIQERLYNRLTLTAISQORVLPNDNCLRPCEMNCVSVL 1247
D 3472 GSARVTPDGLVTRBEGSRBAQEWYQIQASDSGIRPLSLTSVR-----HVT 3522
QY 1248 RFDSSAP-----FIASSVLPFRHPVGLRGCPGFTGDCETEDV--LCYSRP 1296
D 3523 EQSHAPBALPLEIFITVGEDEBQ-----GAWGKIHARDRQDULTYS-- 3567
QY 1297 CGPHRCRSRGGYTCLCRDGTGHEHCVSARSRCCTPGVCNKGCTCVNLLVGFKDCDP 1356
D 3568 -----LAEBETLGNHPSVGAPDG----- 3586
QY 1357 SGDFEKPYCYTTSFPAHSFTFRGAKORHFTLALSFAKERDGLLYNGRNEKDF 1416
D 3587 -----IIAAGL-PRGHT--SFNTVSDGTFTTTAGVAVYVMH 3621
QY 1417 VALFVIOQVOLTFESAGESTTVSPFVPGVSDQOMHTVQKYKPKLLAQGTGLPOGSPSE 1476
D 3622 VGOEALQOAMMMGF-----YQLTPELVSDHMRILQFLSHKIDIKRANHLSLQ 3672
QY 1477 OKVAVVTYDGDCTGYALRFGSVLGNYSCAOCTGCGSKSLDLTGPLLGGVPDLPBSFP 1536
D 3673 PAEAVAGVD-----VLVVF-----BGHSSTFEYFQELAS-IITSAKMEHSVG 3715
QY 1537 VRMRQFVCGMKNLQYDSHIDMADFIANGTVPGCPAKVNCDSNTGNGCTVQNDAP 1596
D 3716 VQMR-----SAMPVPCGPTCOGICN-----TV 3741
QY 1597 SCECPGLFGKSCAOMANPOH-----FIASSLVAMHGLSLPISOPWYSLT 1642
D 3742 HLDPRVCGTYSTARSLITPRHRLQSCSCNGTATRFSGQYVYRA--PAARNMHIF 3798
QY 1643 MFRTRQADGVLLQATTRGRSTITLQRBGHMLSVETGGLQASSSLRPGANGDMH-- 1700
D 3799 YLKTLPQALIL--FTNETASVSLKSLASGVQLBYHCLGSGYVGNLSQ-RHVNDHEWMSI 3855
QY 1701 -----HAQLALASGPGHAILSPDYGOORABENGLPRLHGLHSNITVGG--IPPA 1751
D 3856 LVEEMDAISRLMVMDSMTSLV-----PENCNG-LAPRHH-----LLGGLILHSS 3902

```

```

QY 1752 GGVARGFRCGLQGVRSDFREGVNSLDPHSG-----ESINVEQCSLPDPCDSNCPBAN 1806
D 3903 SNVSQGEGLDAVVNNE--EALDLAPKRTVAGLLETQALTOCCLSHSDYCSQNTCLNG 3960
QY 1807 YCSNMD--SYSCSDPGYGDNC--TNVCDLNFCEHQSVCYTRKPSAPHGYTCBCPPN 1860
D 3961 KCS--WTHGAGYCKCPGPGSGHRCQGRNCTFAFLCBGGTCI--LSPKASCNCFHP 4015
QY 1861 YLPGYCETR-----IDPCPPGWWG 1880
D 4016 YTDRCCEMARGCSBGHCLVTPRIQRDWDG 4045

RESULT 15
FAT2 FAT2_ID FAT2_RAT FAT2_STANDARD FAT2_PRT; 4351 AA.
AC 088277;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (Multiple epidermal growth factor-like domains 1).
GN FAT2 OR MEGF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
CC -1- SIMILARITY: Contains 33 cadherin domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
D EMBL; AB011527; BAA32458.1; -
D PIR; T00252; T00252.
D HSP; P08709; IBF9.
D InterPro; IPR002126; Cadherin.
D InterPro; IPR000742; EGF 2.
D InterPro; IPR006209; EGF-like.
D InterPro; IPR006210; IEGF.
D InterPro; IPR001791; Laminin G.
D Pfam; PF00028; cadherin; 33.
D Pfam; PF00054; Laminin G; 1.
D PRINTS; PR00205; CADHERIN.
D SMART; SM00112; CA; 33.
D SMART; SM00181; EGF; 2.
D SMART; SM00282; Lang; 1.
D PROSITE; PS00232; CADHERIN 1; 14.
D PROSITE; PS50268; CADHERIN 2; 33.
D PROSITE; PS50025; LAM G DOMAIN; 1.
D PROSITE; PS00022; EGF_1; 2.
D PROSITE; PS01186; EGF_2; 2.
D Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
D SIGNAL 18
D CHAIN 19 4351
D DOMAIN 19 4050
D TRANSMEM 4051 4071
D POTENTIAL 4072 4351
D CYTOPLASMIC (POTENTIAL).

```

FT	DOMAIN	34	148	CADHERIN 1.
FT	DOMAIN	149	256	CADHERIN 2.
FT	DOMAIN	363	458	CADHERIN 3.
FT	DOMAIN	459	564	CADHERIN 4.
FT	DOMAIN	565	669	CADHERIN 5.
FT	DOMAIN	716	820	CADHERIN 6.
FT	DOMAIN	821	925	CADHERIN 7.
FT	DOMAIN	926	1032	CADHERIN 8.
FT	DOMAIN	1033	1142	CADHERIN 9.
FT	DOMAIN	1138	1242	CADHERIN 10.
FT	DOMAIN	1243	1346	CADHERIN 11.
FT	DOMAIN	1350	1448	CADHERIN 12.
FT	DOMAIN	1449	1555	CADHERIN 13.
FT	DOMAIN	1556	1660	CADHERIN 14.
FT	DOMAIN	1661	1758	CADHERIN 15.
FT	DOMAIN	1759	1872	CADHERIN 16.
FT	DOMAIN	1873	1968	CADHERIN 17.
FT	DOMAIN	1969	2070	CADHERIN 18.
FT	DOMAIN	2071	2171	CADHERIN 19.
FT	DOMAIN	2172	2272	CADHERIN 20.
FT	DOMAIN	2273	2379	CADHERIN 21.
FT	DOMAIN	2380	2481	CADHERIN 22.
FT	DOMAIN	2482	2585	CADHERIN 23.
FT	DOMAIN	2586	2692	CADHERIN 24.
FT	DOMAIN	2693	2799	CADHERIN 25.
FT	DOMAIN	2800	2908	CADHERIN 26.
FT	DOMAIN	2909	3013	CADHERIN 27.
FT	DOMAIN	3014	3115	CADHERIN 28.
FT	DOMAIN	3116	3220	CADHERIN 29.
FT	DOMAIN	3221	3323	CADHERIN 30.
FT	DOMAIN	3324	3428	CADHERIN 31.
FT	DOMAIN	3429	3533	CADHERIN 32.
FT	DOMAIN	3534	3631	CADHERIN 33.
FT	DOMAIN	3775	3946	LAMININ G-LIKE.
FT	DOMAIN	3949	3986	EGF-LIKE 1.
FT	DOMAIN	3988	4024	EGF-LIKE 2.
FT	DISULFID	3953	3964	BY SIMILARITY.
FT	DISULFID	3958	3974	BY SIMILARITY.
FT	DISULFID	3976	3985	BY SIMILARITY.
FT	DISULFID	3992	4003	BY SIMILARITY.
FT	DISULFID	3997	4012	BY SIMILARITY.
FT	DISULFID	4014	4023	BY SIMILARITY.
FT	CARBOHYD	39	39	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	210	210	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	459	459	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	627	627	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	789	789	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	966	966	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1276	1276	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1417	1417	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1899	1899	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1998	1998	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2007	2007	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2102	2102	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2165	2165	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2183	2183	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2325	2325	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2368	2368	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2387	2387	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2430	2430	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2470	2470	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2547	2547	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2597	2597	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3127	3127	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3278	3278	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3312	3312	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3432	3432	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3603	3603	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3770	3770	N-LINKED (GLCNAC. . .)

FT	CARBOHYD	3774	3774	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3815	3815	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3842	3842	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3875	3875	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3906	3906	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3991	3991	N-LINKED (GLCNAC. . .)
SO	SEQUENCE	4351	AA; 480644	MM; C564E7F98B8388 CRC64;
Query Match				
Best Local Similarity 24.0%; Pred. No. 8.9e-53;				
Matches 534; Conservative 271; Mismatches 830; Indels 588; Gaps 77;				
Qy	179	PQGPSPYQATVPENDPAGTPVASLRAIDPDBGAGLRYMDALFDSRNOFSLDPT	238	
Dy	2377	PKRFPQYKANVSELATGCHLVKQALDPDIDGTSRLER---	LISGNDRHSINSTS	2433
Qy	239	GATYTAELDRERKSTVFRVTADHGMFRSALATITLVTDNHDVPPEQOEYESTL	298	
Dy	2434	GIISMENLCKKQDLSYNLVGASD-GVFR--ATVPYINTTANKKSPERQONVYAEI	2490	
Qy	299	RENLEVEYEVLTVRATPDGAPNANILYRLBSGSGSPSEVFRIIDPRSGVIRTRGPVDR	358	
Dy	2491	AENAKGTKTIELLAIKDSDGPGYGTDTYINTLAG---	ERFPLNR-GQITTIQKIDRE	2546
Qy	359	EVSRYQITVEASDQGRDPGRSTTAAVFLSVEDNDNAPQFSRRYVYQVEDVTPGAPV	418	
Dy	2547	--NSTERVIAIKVMARDGGKVAFCVTKIILTDENDNAPQFKASGYVISIPSNVSDSP	2604	
Qy	419	LRTYASRDGSAVNVHYSTMS-----		440
Dy	2605	IQVLAADADGRNADVYTSVSTEDLABEIIIEVNPPTGVVKEISLVLENRAVDENIKA	2664	
Qy	441	-----		440
Dy	2665	QDGGPRHMDSLVVRQOVNENIPLPKFSRPLYTPFSAPEDLPBGSEIGSKAVAAQDPI	2724	
Qy	441	-----GNARGPYLDAQTGALDVSPLDYETTKETYLVR---	ODGGRPLSN	486
Dy	2725	YSIVQGTTPSPNSDDPFSLDQDTGVKVRKAMHSESTKVIQIDIMACHPREDTDLVLS	2784	
Qy	487	VSGIVTVQVLDINDNAPIFSTPFOATVLESVGLVILVQALIDADGNANLEVLA-	545	
Dy	2785	VS-----IQVEDVDNDRPVEFADPYKAPLTENMGQTTVQVTDNDQTSIDGQVSTRLSV	2840	
Qy	546	--GVGHDPEPTINNGTWSVAALDRREVDYFSPGEARDHG--TPALTASASVTVLD	602	
Dy	2841	EPSPNHELPAVDSESGWITTLQEDCETQYTRFYVAFHDQOTIQLSSQALVVISITD	2900	
Qy	603	VNDNPTFTQPERYTVRLNEDAVGTSVTVSAVDRDA--HSVITYQITSGNTRNRPST	659	
Dy	2901	ENNDPREFASEDYRGSVENNEBGEI.VATLKTLDADVSDQNRQVTCYITGEDPLGQFSS	2960	
Qy	660	SQGGGGLVSLALDLYKLEQVLAATASDGTRODPAQIVVNTDANTHRPVGQSSHYTV	719	
Dy	2961	QVEDMERISRRKTLDBHAKYLLRVTASDGKQOASVPEVAVVDINDNSPQSQLLYTG	3020	
Qy	720	NNEDRPAGTTVVLIISATBEDTGENARITYFMS--DISIPRIDADGAVTQALDYEDQ	778	
Dy	3021	KAEEDVTPGHFILAQVAILVDMDTNAQITYSLHGPAGQEKLDPHRGELTTLVLDRERK	3080	
Qy	779	VSTYLAITARDNGIPKQSDTTYLEILVNDVNDAPQFLRDSYGSVYEDVPFTSVLQIS	838	
Dy	3081	DVNLVAKKADGG--QOSQCAEVTLHIEVDNDNAPRPFPSCHCVAVFDNTTVKTPAVVVF	3138	
Qy	839	ATRDGSLNRVYPTQGGDDGDDIVSTSGIVTTLRLDENVAQVYLRAVADKGM	898	
Dy	3139	ARDPDQANNAQVYSLT--DSADGQFSIDATSSVIRKEKPLQVRASAVELTVRASDLGT	3196	
Qy	899	P-PARTMEVTVLVLDVNDNPPVFEODEPVEFVENSPIGLAARV-TATDP--DEGTNAQ	955	
Dy	3197	PILSLSTIGTVTSVGLLEDVLPFIINAHSSTQVPEBAPIDIMETVHLATILRPSSEKTG--	3254	
Qy	956	IMQIVGNIPEVQDITPSGELTALVLDLYEDRPREYLVIAQ--TSAPLVSRATVHR	1012	

```

Db 3255 --YITGNEGCKRLAHTILYVNGSLPFEINPKYFLSIECRKSSSISDVITVIN 3312
QY 1013 LLDENDNP-----VLGNPEILFNNYVTRSSSPFGAIGRVPAPDP--ISDSLVSFER 1066
Db 3313 VTDVNEHFRTHLTVRVLENAV-----GDVILTVASDDEGPNASAITVSLVG 3364
QY 1067 GNEILSLVLNASTGBLKRALLDNNRPLEAIVSLVSDG---VHSVTAQCALVITITD 1122
Db 3365 GNQGHFTINPKKGLQVAKALDWEQTPSYSLRLRATDSGQPLHEPT-EVAEIVDVND 3423
QY 1123 EMLTHSTLRLEDMSPERFLSLGLPIQAVATLAPPDHVVVFNQORDADAGHIILN 1182
Db 3424 -----NPRFQO-----LMYSTSVQENSPIG-1KV 3447
QY 1183 VLSLVGPBGPGGPPFL-----PSEDLQERLYINRSLTAISAQVLPFDNICLEBPC 1237
Db 3448 LQILIDBDSFONGPFRITBGNISVFRVTDGRLVTAASLS-----KKAR 3496
QY 1238 ENYRCVSVLRFSSAPFIASSVLFREIHVGLRCCRP-----G 1279
Db 3497 EMYOLHIEV--SDSGLPPLSSSTLVVQVTE---QSRYPSTLPLRISITKGEERFQGG 3550
QY 1280 FTGDYCEBYD---LCYSPCGHRCRSGEGYTCLCRDGYTGHCEVSAASGCTPG 1335
Db 3551 MIGIHAIDRDPDITLTS-----LEDEGLDRYFTVAGSDGK--- 3588
QY 1336 VCKKGGTCVNLVGGFKDCPSGDFEKPVCQVTRSPAHSPITFRGL-RORFHTLALS 1394
Db 3589 -----IASQILPHGRYSPNVTVS 3607
QY 1395 FAYTERDGLLYNGRFNEKDFVALLVIOEQVLTFSAGESTTVSPFVPGVSDGQMT 1454
Db 3608 ---DGTFTTGVHVHVMMEBEVPOQAVMLGF-----HOLTPELVSDBWRN 3652
QY 1455 VOLKYKPKLLGQGLPGSPSEQKAAVTVVDCDGTVALRFGSVLGNTSCAOGTQGGSK 1514
Db 3653 LQRLSLNLDVKRANIHILASLOPAEVTAGVD---VLVLF-----ERHSQTS 3695
QY 1515 KSLDLTGFLILGVPDLPESEFVVRKQFVGCMRLQVDSRHIDMADFIANNGTVPGCCPAK 1574
Db 3696 YDLQELASAIHNSVREIHSVIGIR-----SALPVVPCQ 3730
QY 1575 KNVCDNSTCHNGGTCVANOQMDAPSCCEPLGFGGKSCAOEMANPOHFLGSSLVAMHGLSLPI 1634
Db 3731 GQSCODQTCQ-----ETVLSLEPRVGPSTVARSILTPRHILGRN-CSCNGTTLRF 3780
QY 1635 S-----QPMYLSLMFRTRQADGVLLQAITRGRSTITLQREG-HVMLSVEGT 1680
Db 3781 SGQSYVQYRPLEADQWQIHFYIKTLQPMALLM--FTNETASISLKLANGFSLHYHCPG- 3837
QY 1681 GLQASLSLLEFRANDGDHMHQAOLALGASGFGHAILSPDYGOQARAEGNLGPRLGLHLS 1740
Db 3838 GFYGNLSRYP--VNDGGMH--SMLEBRDTSVHLVDI---TDNLSLVIPEBCGLRTE 3890
QY 1741 -NITVGGI--DPPAGGVARGRGCLQGYRVG-DTPEGVNSLDPHSG--ESINVEQCCSLP 1794
Db 3891 RQLLIGLIVPNSPSSNVLGFEGCLDAVAVNGERLELLGREKKMEGRLETWALSQCCWPG 3950
QY 1795 DPCDSNPPCANSYCSNDMS--YSGSCDBGYGDNC---TNVCDLNPCEHOSVCTRKPAP 1850
Db 3951 TACGSPCLNGSGCPALGSGYLCCPPFSGRNCLEGRNCTSAPOEGGTCV---SSP 4007
QY 1851 HGYTECEPNTYLPYCETR-----IDOPCPRGMWH-----PTCG 1885
Db 4008 EGTSCNCFHPTGDRCEMBARGCSGHCILITPEIKRGDMGOQEFVITVALPLVITATVG 4067
QY 1886 -----PCNCVSKGDPDPCAKTS-----GECHCKENHYRP 1915
Db 4068 LLLYCRRRKSHKPYTME---DPLLARISIGVDTQASPAIELDPLNTSSCN---NLNQP 4119
QY 1916 PGSPTCLDCDCTPTGSLR---YCDPBDGQPC---KPGVIGRQCDRCNPFABVTTN 1967

```

```

Db 4120 EPEKTSVPELVTGSSKQRPWVCVPPRLPPAAVSSHG-----HEPIIKRTWS 4170
QY 1968 GCEVNTDSCPRATEAGIM-----W-----PRTFRGLPAAAPCPKGS 2003
Db 4171 GEBLIVPS-----GAAVPPTYSRKGMEXPHPEMTQGLTPSPRHHVG-PAVMPDPTGL 4224
QY 2004 FG--TAVRHCDERHMLPPNLFN-----C-----TSITF----- 2030
Db 4225 YGCFPEPLENKGRLPPRYSNQNLDMPPRPPSPREHLAPCLMEYTAISYHSQFR 4284
QY 2031 -----SELKGFARLQRI-----ESGLDSGRSQQLALLNNAQHTAGYFG 2071
Db 4285 QGGGSPCLAEGGYKGVSMLSRAGPSYADCEVNGGATGRSQPRA-----PNYEG 4335
QY 2072 SDV 2074
Db 4336 SDM 4338

```

Search completed: February 11, 2004, 15:55:45  
 Job time : 49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:52:09 ; Search time 76 Seconds  
(without alignments)  
9924.839 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545  
Sequence: 1 MESPATGVPLPTPPPLLL.....ACTVDESSGSEPLFNFLLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_rvirs:\*  
16: sp\_Bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5065.5	32.6	1766	13 Q8AW45	Q8AW45 brachydanio
2	2491	16.0	2610	5 Q19482	Q19482 caenorhabdi
3	1538	9.9	2909	5 Q8MY78	Q8MY78 asserina pe
4	1520	9.8	4589	11 Q9WU10	Q9WU10 rattus norv
5	1519	9.8	4610	13 Q8AXB7	Q8AXB7 brachydanio
6	1508.5	9.7	2809	5 Q61230	Q61230 lyechinus
7	1496	9.6	4587	11 Q9QX3	Q9QX3 mus musculu
8	1413	9.1	4555	11 Q8R508	Q8R508 rattus norv
9	1164.5	7.5	2920	5 Q967F4	Q967F4 caenorhabdi
10	1108.5	7.1	1959	5 Q8MY75	Q8MY75 ptychodera
11	1092	7.0	4307	5 Q19319	Q19319 caenorhabdi
12	1081.5	7.0	1061	4 Q8NSB3	Q8NSB3 homo sapien
13	906	5.8	1881	5 Q8T5J8	Q8T5J8 anopheles 9
14	905.5	5.8	803	4 Q96JL3	Q96JL3 homo sapien
15	905	5.8	1820	5 Q9YVGO	Q9YVGO drosophila
16	905	5.8	1820	5 Q8IGX4	Q8IGX4 drosophila

17	862.5	5.5	2005	5 Q9VFH5	Q9VFH5 drosophila
18	853	5.5	1060	4 Q8IUP2	Q8IUP2 homo sapien
19	850	5.5	1037	4 Q9H4E1	Q9H4E1 homo sapien
20	850	5.5	1037	4 Q8MY34	Q8MY34 homo sapien
21	850	5.5	1048	4 Q9BZA9	Q9BZA9 homo sapien
22	850	5.5	1340	4 Q9BZA8	Q9BZA8 homo sapien
23	848	5.5	1038	11 Q8CFX3	Q8CFX3 mus musculu
24	847	5.4	1021	4 Q9P2X5	Q9P2X5 homo sapien
25	847	5.4	1025	4 Q9H4E0	Q9H4E0 homo sapien
26	847	5.4	1337	4 Q9BZA6	Q9BZA6 homo sapien
27	847	5.4	1347	4 Q9BZA7	Q9BZA7 homo sapien
28	833	5.4	1310	4 Q96RV4	Q96RV4 homo sapien
29	833	5.4	1347	4 Q96RW0	Q96RW0 homo sapien
30	831.5	5.3	965	11 Q8QZT6	Q8QZT6 mus musculu
31	814.5	5.2	961	4 Q8ND89	Q8ND89 homo sapien
32	809.5	5.2	1032	6 Q95KD8	Q95KD8 macaca fasc
33	809.5	5.2	1203	4 Q9HC56	Q9HC56 homo sapien
34	786.5	5.1	1016	13 Q9QZT2	Q9QZT2 xenopus lae
35	780.5	5.0	1035	13 Q57537	Q57537 xenopus lae
36	768.5	4.9	1069	11 Q88185	Q88185 mus musculu
37	753	4.8	1563	5 Q9VAF5	Q9VAF5 drosophila
38	739	4.8	733	6 Q95KK0	Q95KK0 macaca fasc
39	738	4.7	1084	4 Q8N3K7	Q8N3K7 homo sapien
40	735.5	4.7	1469	4 Q9HAK3	Q9HAK3 homo sapien
41	735.5	4.7	1474	4 Q94910	Q94910 homo sapien
42	733.5	4.7	1030	13 Q9QZT4	Q9QZT4 gallus gall
43	729.5	4.7	1467	6 Q97830	Q97830 bos taurus
44	729.5	4.7	1472	6 Q97831	Q97831 bos taurus
45	728	4.7	1515	11 Q88917	Q88917 rattus norv

## ALIGNMENTS

### RESULT 1

ID	Q8AW45	PRELIMINARY;	PRT;	1766 AA.
AC	Q8AW45;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	SI:BZ6108.1 (Novel protein similar to cadherin, KDP IAG seven-pass G-type receptor (CRLSR)) (Fragment).			
GN	SI:BZ6108.1.			
OS	Brachydanio rerio (zebrafish) (Danio rerio).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Osteichthyes; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.			
OC	NCBI_TaxID=7955;			
OX	NCBI_TaxID=7955;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Babbage A.;			
RL	Submitted (DRC-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL645790; CAD58738.1; -.			
FT	NON TER			
FT	SEQUENCE 1766 AA; 197485 MW; 80CAAA18F5A497F CRC64;			

Query Match	32.6%	Score 5065.5;	DB 13;	Length 1766;
Best Local Similarity	52.3%	Pred. No. 8.5e-312;		
Matches 976;	Conservative 279;	Mismatches 431;	Indels 181;	Gaps 37;
QY	1105 GVHSVTAQALRVITIDEMITHTSLRLSDMSPERFLSPILGLFIQVNAATLTPPDHV	1164		
DB	1 GHQVTAFCRLRVITIDEMITHTSLRLSDMSPERFLSPILGLFIQVNAATLTPPDHV	60		
QY	1165 VFWVQDPTDAPGHHIIVNSISVQPPGPGGPP--FLPSRDLQERLYIKNSLITAIASQ	1222		
DB	61 FIFVQNDTV--SGNINLVFSALP---GGVDRYFPSEHQEQIYIKNTLLQKISSQ	115		
QY	1223 RVLPFDNIIICREPCENYKRCVSLRFPSSAPFLASSSVFRRPIHPGGLRCRCPPPTG	1282		
DB	116 NVLPFDNIIICREPCENYKRCVSLRFPSSAPFLASSSVFRRPIHPGGLRCRCPPPTG	169		

QY 1283 DYCETEVLDLCYRSPGPHGRCSREGVYLCRDGYTGEHCESVARSGRCTPGVCKXNGT 1342  
 DB 170 -----ENCEVDSRGRCVPGCKNGE 191  
 QY 1343 CYNLLVGGFKDCDSCSGDEPEKYCOVTRTSPRAHSFTRGRLRQNFHTLALSPATKRDG 1402  
 DB 192 CYNLLVGGFKDCDSCSGDEPEKYCOVTRTSPRAHSFTRGRLRQNFHTLALSPATKRDG 251  
 QY 1403 LLLYNGRPNRKHDPVLALEVIQEOYOLFPSAGESITTVSPFVPGVSDGQMTVYOLKYYNK 1462  
 DB 252 LLLYNGRPNRKHDPVLALEVIQEOYOLFPSAGESITTVSPFVPGVSDGQMTVYOLKYYNK 311  
 QY 1463 PLLGQTLPGSPBQKXAVVTVDCDGTVALRFGSVLGANSYCAAQGTQSGSKSLDTLGP 1522  
 DB 312 PKISHLGLPGSPBQKXAVVTVDCDGTVALRFGSVLGANSYCAAQGTQSGSKSLDTLGP 371  
 QY 1523 LLLGCVDPDLPSPFVPRKQPFVGCNRNLOVDSRHIDMADPIANNGTVECPAKKXVCDST 1582  
 DB 372 LLLGCVDPDLPSPFVPRKQPFVGCNRNLOVDSRHIDMADPIANNGTVECPAKKXVCDST 414  
 QY 1583 CHNGGTQVQNDPASCCEPLFGGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPTLSTL 1642  
 DB 415 COHAQENCKNNTKPCCEBGRGKXKCDQMPKOHFDGHAMMLNDPMTIAVPMYIAL 474  
 QY 1643 MFRTRQAD--GVLLQAITRGHSTITTLQREGHVMSVEGTGLQASSLRLEPGRANDGDWH 1700  
 DB 475 MFRTRQAD--GVLLQAITRGHSTITTLQREGHVMSVEGTGLQASSLRLEPGRANDGDWH 519  
 QY 1701 HAQALASAGS-----PCHALISPDVGOQRAEGNLPRLHGLHSLNITVGITGPRAGV 1754  
 DB 520 HLVLLESSSKDGKDKTKMAOVLIDYDMFKSVE--IGNELPGLTKKSPFIOGLOGODIV 577  
 QY 1755 ARGGRGLQGVAVSDTEGVNSLDPSSHGESINVEQSGSLPDCPSNCPANSYCSNMDS 1814  
 DB 578 QGGKGCQGGKARMETATSTANIMHHAQKTRVEDGCDMSANCALINCEKSOCT--De 634  
 QY 1815 YSCSGDPGYGDNCTNVCNLPCEHQSVCYTRKPSAPHGTYCECPNLYAGPYCETRIDQPC 1874  
 DB 635 HTCTCDPFGFGRDVCDAChLNPCEHLSTCYRKPSSSHGTYCECQDYGYGCEKVKPC 694  
 QY 1875 PRGMWGHPTGCPNCDVSKGPRDCKNTSGEHCCKENHPRPGSPCLLDCYRTGSLR 1934  
 DB 695 PRGMWGHPTGCPNCDVSKGPRDCKNTSGEHCCKENHPRPGSPCLLDCYRTGSLR 754  
 QY 1935 VCDPEBQCPCKPVIIGRQCDRCNPNFALEVTTNGCEVNYDSCPRALBAGIWMPTRRGLP 1994  
 DB 755 TCDBETQCPCKKAVIGRQCDRCNPNFALEVTTNGCEVNYDSCPRALBAGIWMPTRRGLP 814  
 QY 1995 AAAPCPKSGFPTAVRHCDEHKGWLPNLFNCTSTPSELKGFARLQRNESGLDSGRSQ 2054  
 DB 815 VAMNCPKSGISTAVRHCDEHKGWLPNLFNCTSTPSELKGFARLQRNESGLDSGRSQ 874  
 QY 2055 LALLRYAHTOHTAAGVPSDVVAQOLATRLAHSTQGRGLSTQVHPTRENLARTGSA 2114  
 DB 875 IARLRYAHTOHTAAGVPSDVVAQOLATRLAHSTQGRGLSTQVHPTRENLARTGSA 934  
 QY 2115 LLDLTANRAMELIQOTEGTAMLIQHYEAYASALQONRHTYLSPTITVENIYISVRL 2174  
 DB 935 ILDSNKEHMOIQTREGTANLKHREYANTLAQONRHTYLSPTITVENIYISVRL 994  
 QY 2175 ----DKGNFAGAKLPRYEALRGEPDLETTVILPESVRETPEVVPAGGEAQBEEL 2230  
 DB 995 VTYPD-----ATKYNFQTDQCECPKDKSSVLPPEFESFKSHKODPEFOINDEBEKO 1049  
 QY 2231 A-RBRRHPELSQGEBAVASVITYRTAGLPHYNDPDRKSLRVKRPINPVYSIVHD 2289  
 DB 1050 ATKRRKRVETIPPEPVASVITYRTAGLPHYNDPDRKSLRVKRPINPVYSIVHD 1109  
 QY 2290 DEBLPRALDPVTVQFLLETBERKTEICVFMNHSILVSGTGMASRGCEVFRNESHV 2349  
 DB 1110 EGGLHPIEREPITLDVNLLETBERKTEICVFMNHSILVSGTGMASRGCEVFRNESHV 1169  
 QY 2350 SCQCHMTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTPFFLTLRLIRSNQ 2409

DB 1170 SCQCHMTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTPFFLTLRLIRSNQ 1229  
 QY 2410 GIRRNITAGLQVFLGINDADIPACTVAILLHPLYLCTFSMALLBHLRYALT 2469  
 DB 1230 SHKNVVALFSELEVLGINDADIPACTVAILLHPLYLCTFSMALLBHLRYALT 1289  
 QY 2470 EVRDVNTGPMRPFYVYMGVGPATITGLAVGLDEBGVGNPFCVLSYDITLINSFAGVAF 2529  
 DB 1290 EVRNINHGMRPFYVYMGVGPATITGLAVGLDEBGVGNPFCVLSYDITLINSFAGVAF 1349  
 QY 2530 AVMSVFLYTLAARASCAAROGFEKKGPVSLQPSFAVLLLSATWLALLSVNSDTLL 2589  
 DB 1350 VLVNIVIFWMAKASCGRRQSEKSGVTPALRMAFLLLLSATWLGLMAVNSDVT 1409  
 QY 2590 FHTLPATCNCIOGPFITLVSYSKVRALK-LASRKPSPPALITST-LTSSYCP 2647  
 DB 1410 FHTLPATCNCIOGPFITLVSYSKVRALK-LASRKPSPPALITST-LTSSYCP 1469  
 QY 2648 SPY-ADGRLYQ--PYGSPAGSLHSTRSGKSPSYIFPLRBSALNP--GQSPGLGD- 2701  
 DB 1470 NTMEDGALFRALIGESTYSLESTVSAKSHSYFTYTRDELQCPVSQNAKAGITDI 1529  
 QY 2702 PGSLFLEGDDQHDPTDSDSLSLBDQSGVASTHSSDSEEBEBAAPFEGQW 2761  
 DB 1530 DGSILFRNGTAD--DSDSDSLSV-DEHSSYASHSSDSEEDIDMOPK-----W 1577  
 QY 2762 DSLGAGARLPKHSHPK-----DGGPBGKAWPDPFGTAKSSGNGAPERLRENG 2816  
 DB 1578 NN-----EKOPHLSHPKTEVDVTSNHYKPPWPTAUNTASDSEDPGA--ERLVEYK 1629  
 QY 2817 A--LSREGSLG-----PLP-----GSSAOP--HKGLKKCL--PTISEKS--S 2852  
 DB 1630 VNTLHGEKMLNIGESTQDKQPTPSIQANSNHOEGKGLIKNTITTPPTIDGMCK 1689  
 QY 2853 LIRLPLEQC--TGSSRGSS--ASEGSRG--PPRP--PPROSLQEOQNCVPMIAMS 2901  
 DB 1690 KIREKLSVYMPPTISSRAVITSNDGANGHVLKIPRPVPPR-----EHLANG--MAN 1742  
 QY 2902 IKAGTVD 2908  
 DB 1743 LKSAVVA 1749  
 RESULT 2  
 ID 019482 PRELIMINARY; PRT: 2610 AA.  
 AC 019482; Q23218;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE P15B9.7 protein.  
 GN P15B9.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Baynes C.;  
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
 RX MEDLINE=94150718; PubMed=7906398;  
 SEQUENCE FROM N.A.  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,  
 RA Litching J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Riken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL elegans.";  
 RN Nature 368:32-38 (1994).  
 RP [3]  
 SEQUENCE FROM N.A.  
 RA Percy C.;  
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; Z78018; CAB01449.1; -;  
 DR EMBL; Z78013; CAB01449.1; JOINED.  
 DR EMBL; Z78013; CAB01427.1; -;  
 DR EMBL; Z78018; CAB01427.1; JOINED.  
 DR HSSP; P15116; INCI.  
 DR WormPep; F15B9.7; CE09412.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006209; EGF\_Like.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR001791; laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF01825; GFS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 8.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00303; GFS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 6.  
 DR PROSITE; PS0266; CADHERIN\_2; 8.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS0116; EGF\_2; 2.  
 DR PROSITE; PS0221; GFS; 1.  
 DR PROSITE; PS0227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS0025; LAM\_G\_DOMAIN; 2.  
 DR KMW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KM Glycoprotein; Repeat; Transmembrane; EGF-like domain;  
 SQ SEQUENCE 2610 AA; 290075 MW; 24184B4C836805CA CRC64;

Query Match 16.0%; Score 2491; DB 5; Length 2610;  
 Best Local Similarity 27.4%; Pred. No. 4e-146;  
 Matches 768; Conservative 438; Mismatches 1205; Indels 394; Gaps 89;

QY 51 CAPMGLCPSSASNLMLYTSR--CRDAGTELGHVPHHDGLRWCPESEAH1PLPPAP 108  
 DB 39 CRKCA--VPSNSNVIMLPASRPCLHPGQPII--HMPDISDMLA-----CPVGLPD 86  
 QY 109 GCMWCRLLGIGHLSPQGLTLPEHPC-LKAP-----RLRQSCSLADAPGL 156  
 DB 87 SVHSSQISL-----LRGEG-LLTKEKICFPDGPIDFHYDYCDGLKYSKMTIGHSIAS 140  
 QY 157 RAGERSPEESIGGRKRKNVNTAPQPPSYQATVPENQAPGTVPASLRATIDPGRAGRL 216  
 DB 141 KKLLEFRRTKRWARRRNPDAVHFOEKVYKELPBTPLETIIASVKA---SHASSQPL 197  
 QY 217 EYTMALPDSRSNQFSLDPVTGAVTTAELEDRKTSVHFVYTAQDHGPRRSALATLT 276  
 DB 198 YYSMVAPQDSRSQNLFTLDTMSGEIRLAKMDREVDKHLIKTAYAEVDPFTISASTTV 257  
 QY 277 ILVTDNDHDPVEQGEYKSLRENVGEVYLVATDGDAPNNMLLYRLLEGGSGS 336  
 DB 258 VHVLDVODNSPIREKSYFGEIREDAPIGTVLVSPARDSDSGENGEIETSLGEGNG--- 314  
 QY 337 SSVFEIDPRSGVLRTRGVPVREKVESYQLTVBASDQGRDPGRSTTAAVFLSYEDNDNA 396  
 DB 315 KNLAINAKSGVGTAPALDRRTLSLRDLVIASDKGP--KRESTAMWEIIVDVYNDNA 372  
 QY 397 PPSSEKRYVQVEDVTPGARVLRVTAASDDKSNVAVHYSNGARQGFYIDAQTGL 456  
 DB 397 PPSSEKRYVQVEDVTPGARVLRVTAASDDKSNVAVHYSNGARQGFYIDAQTGL 456

DB 373 PVFASDYNVTLLENITTPAVIATVKATDEDEFTNGKHYVSNASSSGIGLTIIDYSTGEV 432  
 QY 457 DVVSPLDVETTKETYLARAQDGRPLSNVSGLVTVQVLDINDNAPIVSPFPQATVLE 516  
 DB 433 TLREIRIDAKNS-PITAVIRAKDQAQPALSTVPL-TINVIDINDAPITLIAQKMTLEE 490  
 QY 517 SVPLGYLVHQAIDADAGDNARLEYLAGVGHDPPTINNCGMIVSAAEIDREEDVY 576  
 DB 491 NVAIGBEGRVVAIDEDSGPNIGIKYSMEG---SDFIIDESGILKTKLDRRTTAY 547  
 QY 577 SGVEARDHGIPALPASASVTVLDVNDNFTFPQPEYTVRLNEDAVGVSVTVSAVD 636  
 DB 548 SLKVTARMDGTPSLTSTTIAVVLKDINDNAPTFKKEYNVITISEMPGSOITLKAVD 607  
 QY 637 RDAHVITVQITSGNTRRFSITSSGGGLVSLALPDVKERO---VLTANTASD-GTR 692  
 DB 608 NDEQKITRIEAD-REVSFLIDIGDQAI--LSVSELRQDHKRVETISATDQGL 663  
 QY 693 QDTAQIVNAVTDANTHRPVQSSHYTVVNNEDRPAGTTVLIS----- 735  
 DB 664 QGRCVNVPIIDVNS-APYFNHPPSVKIPHSPIGYPITLKVSPFSYFGYGLKHVY 722  
 QY 736 -----ATDEDTGENARITYFMEDSIPOFRIDATGAVTQAELDYEDCV 779  
 DB 723 SGNLRKKRIPNSANMHAEDHNRGNARIVYSI-DSQEFRIIDPSSGDISVSSDLREDRA 781  
 QY 780 SYTLATITARDNGIPKSDPTYLELLVNDVNDAPQFLDSYOGSYVEDPPTSYLQISA 839  
 DB 782 TFSVIVTASDHAASPPLANTSTQLEVLIDINDNSPQTSSTVATISEDIPVGTSPQVSA 841  
 QY 840 TDRDSGLNGRVYTFQGGDGDGD--FIVESTSGIVRTLRLRDRNVAQVYLAAYADK 897  
 DB 842 IDADIGNGIVDYFLNESSSSPSIQLFLDRISGTLNRSKLDKQGFAYIVLPIARDRG 901  
 QY 896 MPAPATPEVITYVLDVNDNPPVFEQDFVFEENSPIGLAAVAVATADPEGTNAQIM 957  
 DB 902 TFSLSAASEITLTLSDVNDNAPTFEQLSDYLIAENSFGVSTGTVIARADDEGNADIS 961  
 QY 956 YQIVSGNIPVYQQLDIFSGE-----LTALVDYEDDRP-EYVLVQATSAPLVSATVYV 1011  
 DB 962 FRIFGADAKLPIEDDBQNGVVRILTRAEVDYAKANKPFPELOASGQSSVTVPAVI 1021  
 QY 1012 RLIDRNDNPPVLAGNEILLFNNTVTRSSFPGALIGRVAPHDPSIDSLTYSFERGENELS 1071  
 DB 1022 HVSVDNNDKPKADPVILMNRDNYQMAR---QIGFIPAPPDQNAITLEYLE---END 1074  
 QY 1072 LVYLAASGELKLSALDNNRPLBAIMSVSDGVHVSYTAQCALVTTITDMLTHSTLT 1131  
 DB 1075 LIEAKRYTGKILVQKQEWGRN--MDVSPKCVSDGANTECSTCRFLHVLVEPEMLESFTL 1132  
 QY 1132 RLEDMSPERPLSLGLFLQAVATLAT--PRPHVVVFNQORTDAPGHIILNVSLVQGP 1190  
 DB 1133 SLARMTVDQFMDPLVQFRDAMSTLSNWKPSDHIIVGKQHLD--DVIYINAIATDHGR 1190  
 QY 1191 PGFVGGPPLPSEEDLOERLYLNRSLTIAISAQVLPFPDNLICLAREPCENYMCVSLARP 1250  
 DB 1191 VVRGMRATELVESIKK---LEKMTLLQVEYIR-----DESGANPSCSMACRQTKQKV 1242  
 QY 1251 GSAPFIASSSVLPRIHPVGLKRCPPGFT-----GDYCEYVDLCISRPCGHRGRCS 1305  
 DB 1243 GEMKAHEIDNFAKTATVNTVTVCECPGSGFSSGAHGD-CDTRIDECYRGRSNNSTCYA 1301  
 QY 1306 REGGYTCCRGDYGTEHCEVSARSRCIPGVCKNGSTCVNLLVGAFKC-DCP-SGDPEKP 1363  
 DB 1302 FENYTOCECKPFWIRGHCEISYHALTCVPGYMSISLC-ELDGNQMKCHCKHGEDIDE 1360  
 QY 1364 YQVYTRSEPAHSFTTFRGLRQRFATLSPATKBRDGLLLYNRFPNEKHDVFALEYVQ 1423  
 DB 1361 RRLRSVSPDGGGLNVLNLDLPRQTQMTKFRSTIAHGVLVFTG--DRSRDFEVSVDV 1418  
 QY 1424 EGYQLTFSG--ESTTYSPPVPGVSDGQHTVQKTYNKLDAQTGLPQSPSEQKAV 1481  
 DB 1419 RVLKVOFSLGGEKIDARKMENDVENRINDGEMHTALVYSNKOI-----T 1462

QY 1482 VYDGGCDTGYALRFGSVLGNVSCAAGTGGGSKS-----LDLPGELLGGVPL 1531  
 1463 MSIDDCETNSLANT---SPNCIAIKAKALEKCEPTPTPCRYLIDISNGLEFGGRPT 1519  
 QY 1532 PESPPVPRQPFVGCNRLQVDSRHIDMADF--IANNQTV-PGCPAKXVCDN--TCHNG 1586  
 1520 SKQIE---KAFSGCISDLSDYDKEVDPSITKEMHKVQVHEGCKHRDPCSTSDGQSAT 1576  
 QY 1587 GTCNQMDAPSCBCEPLDGFQSKSCQEMANQHLGSLVAMHGLSL---PIQOPWLSL 1642  
 1577 SKCNMRGRI CSCPCSVHSTGECVGLGTDLRGHSLFEESEFVLQPSQVSPFEVSP 1636  
 QY 1643 MPTROADGVLLQAITRGSTITLQREHVMLSV-EGT---GLQASLRLBGRANDGD 1698  
 1637 EFTSRAD---MQVF-----ALBPTQSVHYNLEVDGTILKINIGBSEVELPAPVTSKI 1688  
 QY 1699 WHHAQLAGSGGPGHAILSPDYGOQRAEENLGRHLGLSNITWGGIFGPAAGVARGF 1758  
 1689 WNVVVIKFEAD-----SVATISINGIYSAEAKAS--ISDMNLESLYFGIAGT--GHPSRF 1739  
 QY 1759 RGCLQGRVSDTPEGVNSLDPSHGESINVGQ-----GSLDPDPCD-SNFCPANSYCSND 1811  
 1740 EGCIRNLVD-----GRSISVKKKGKTRAGCVVPMKCSYDSICPABS----- 1781  
 QY 1812 WDSYSCSDPGYVGDNCTNVCYL-NPCEHOSVCTRKPSADPHGYTCECPRYVLPGYCETRI 1870  
 1782 -----DITLPCSVANVCSGTCVSNNTA--GYBICIPAKTKNQLEA 1825  
 QY 1871 -DQCPRGWNG-HPTGCPNCNDVSKGDPDPCNTSGECHKENHYRPPSGPTCLDCY 1928  
 1826 PKQMCPSGWMGTPEPRCRRCSCAOTKDYEAQCDKKTGACQCKKSIFSTING--CVKCEC-G 1882  
 QY 1929 TGSLSRVCDEPDGQCPKPGVIGQRCR---NMPFEVNTNGCEVNVYSCPAIRBGW 1985  
 1883 FGADSTECG-ADGCKCKNGAVGRCDRCRFBQDLSTLK-CRPVSGKCPSEIETSIQ 1940  
 QY 1986 WPTRFGLPAAAPCPKPSFGTAVHCDERHGMPLPPLNCTSIYTFSELKFAERLOPMS 2045  
 1941 WPAQKQSIYRQSCPVGESGLATRKCLJTGMRSDVAMNCTRPVSIWVKKFBLR----- 1996  
 QY 2046 GLDSGSGQQLALIRNATQTAGYFGSDVAVYQALATRLAHSTQRFSLATQVHYT 2105  
 1997 -----PSKILITVANTNTSSIRGNQOIAEALSRLDVYRQSPMKRAHIKMKPT 2050  
 QY 2106 ENLLR-VGSALDPTANKHWEILQCTEGGTAMLLOHTEAVALAONMRTYVSPPTIV 2164  
 2051 EKLIESIGRWSEOPADEYISTLSK-----LNNYAEVVAEIHENV--NPLSPFFVAN 2100  
 QY 2165 PNIVISVRLDKGNFAGAKLPRYEALRGEOPDLETVILPESVRETPEVVRPAGGEA 2224  
 2101 DHIYFADKLDPGNI---LPKFNNF-----VDLRPTGF-----PRVAIYVGT 2141  
 QY 2225 QEBEELARORRHRELSQGEAVASVITYRTLGLILPNYDPDKRSIAVPRKPIINPVVS 2284  
 2142 Q-----VVYSIVPY-----PRCNCNCPMIA 2162  
 QY 2285 ISVADDEELPALDKPVYQFRLLEBERRKPCVPMNSILVSGNSARGCEVFR 2344  
 2163 IVANTSD-----PVIYFEIIEBDGMYTEPCVAFDK-----SGTWTAGALLIGL 2208  
 QY 2345 NESHSQCQNMHTSFAVLMDVSRRENGEILPLKLTLYVAL-GVTLAALLTFFELLART 2403  
 2209 NLTHACEYRIRIGVFTFVNDQSSIVRAQMDMTSPALAGVALFLCFLSILLTSSRS 2268  
 QY 2404 LRNSHGIRRNLTALGLAQLVPL--LGINQADLPFCYVIALIHLHLYICTSSWALTEA 2461  
 2269 LKTHSVRIGLIFPALNILLFVHKTAINDA---YCPRANMLSTSSAPPAWFLYG 2324  
 QY 2462 LHLVRLTEVDVNTGPMRFYMLGMVPAFITGLAVG-----LDPEGGNDPFCWLSI 2515  
 2325 LYIYRMLAD-----GSSSPBLTSSILVGIYVPCILISTTTFVTVDCSLSP 2369

QY 2516 YDLTWSFAPVAFVAVMSVFLYIILARASCAAROGFEKKPGVSGLQSPAVLLLSAT 2575  
 Db 2370 HMLFWCILLPGLPLLSFYAASVLSLKKKDVFAKXNVRAVQHILITFTIG 2429  
 QY 2576 WLLALSVSDTLFHYLPATNCIQGPFIPLSYVLSKVRKALKACSRPS---PD 2631  
 Db 2430 MTLTGLFANQLPLPMEIMESOSII---YLIALVFLWCVDITTKADSNPSMWLDNQ 2486  
 QY 2632 PALTTSTLSSYNGSPYADR-----LVQPYDSAGSLHSTRSQSGOP----- 2677  
 Db 2487 KSMWSESTWADP-QCASPLSPRHQHEVPMDSWVPDVPNSNTH-ITINEBTPRLL 2544  
 QY 2678 -----SYLPFLRESALNPQGPPLGDPGSL--FLBSQDOQHD 2715  
 Db 2545 LQGNRVINILSSPQILNIEGVGHYRNMGSLPLRSNQDEAD 2589

RESULT 3  
 ID 08MY78 PRELIMINARY; PRT; 2909 AA.  
 AC 08MY78;  
 DT 01-OCT-2002 (TREMblrel. 22, Created)  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Ap-cadherin.  
 GN APCAD.  
 OS Aserina pectinifera (Starfish).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;  
 OC Asteroidea; Valvatacea; Valvatida; Aserinidae; Aserina.  
 OC NCBI\_TaxId=7594;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Satch N., Humphreys T.,  
 Zhang S., Tsukita S.;  
 RT "A novel amphioxus cadherin that localizes to epithelial adherens  
 junctions has an unusual domain organization with implications for  
 chordate phylogeny.";  
 RL Evol. Dev. 0:0-0(2002).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; AB075365; BAC06834.1; -;  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001791; laminin\_G.  
 DR Pfam; PR00028; cadherin; 15.  
 DR Pfam; PR01049; Cadherin\_C\_term; 1.  
 DR Pfam; PR00008; EGF; 4.  
 DR Pfam; PR00054; laminin\_G; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 17.  
 DR SMART; SM00181; EGF; 4.  
 DR SMART; SM00179; EGF\_CA; 4.  
 DR SMART; SM00282; LamG; 2.  
 DR PROSITE; PS00010; ASX HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS00268; CADHERIN\_2; 17.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein.  
 SQ SEQUENCE 2909 AA; 316542 MW; 2466441BC3619CF CRC64;

Query Match 9.9%; Score 1538; DB 5; Length 2909;  
 Best Local Similarity 29.1%; Pred. No. 1.4e-87;

Matches 525; Conservative 235; Mismatches 792; Indels 254; Gaps 67;

182 OPSVQATYBE-----NCPAGTPVASLRADDEDEBAGRLRYTMDALFDSRNGPF 232  
 1050 QPPEWSTPDDITWMENISINDVIGILOQOSRS-----GNGVYRVVQOIQIPTMSBEGF 1104  
 233 SL---DPVTGAVTTAEELDRETKSTHVRVTAQDHGMPRRSALATLTLLVTDYNDHDPV 288  
 1105 ANQWVDSMIANLYISAPLDYETTKFYTLQVEAYDDG-NRLSTLIGQRVVNDQVNDDETQ 1163  
 289 PQQOEKESLRLELVEGYEVLTATDGDAPRN-ANILRLRLSGSGSSSEVFEIDPRG 347  
 1164 FVYTFEFAFPEVDNAPFLVTRVQADADTETAFTKITYSL--DQTFSAQYFRIDPPTTG 1221  
 348 VTRTGSPVPREVEESYQLTVEASDQGRDP-----GPRSTTAAFVLSVEDNDNNAPOFSBK 402  
 1222 DLSTTKFPRSESDTSVQVIATDGA PSSLRAGSGNKGYLKVINVDKNDPFPDRP 1281  
 403 RYVVOVREDVTPGAPVLRVATSDRDKSNVAVHYSIMSGNARQOFTLDAQTGALDVVSP 462  
 1282 MYTRTIRREDEBPVGYEVIGVATATVDPS--IPRYLTANNNGAFEVDPATGALSIASPL 1339  
 463 DVETTKETLRYRAOQGRPELNSVGLVTVQVLDINDNAPITVSTFPQATYLESP-IG 521  
 1340 DYEIOREYVWLYSANDG---LNVASTVRIQLENNDEKPEFELPVYTAEYSENDPNVP 1395  
 522 YLVAVQALDADA-GDNARLEYRLAGVGHDPFTINNGTGMISVAELDREBYDFSPG 580  
 1396 ROLQVTAVDGDADVAQDIRSLBGTGCTTTIDAATGQITLTMBLRESVSITKLIIV 1455  
 581 EARDHG--TPALTASASVSVTVLDVNDNPTFTQPEYTVRLMEDAVGTSVVTVSANBD 638  
 1456 KATDGGGTGTSITGVADLEVTVLDENDGPFPELETRGSPVENSRCGTFVWQVTAIDYD 1515  
 639 ---AASVITYQ-ITSGNTN---RSTISQSGGGLVSLALPLDITKLEQYLAATASG 690  
 1516 DNQGRITYSTISSEIPNDSSGLFQIDSGTGMVTVKSGAVLDREKNDYTLRLIRASPV 1575  
 691 TRQDT-AQIVWVNTDANTHRPVFQSSHYTVNNEDEPAGTVVLSATDEDTGENARITY 749  
 1576 PSEALTDIAIQLDVNDNAPVFGGPIATSEBQPVATWERSYTDADVDFBEVOP 1635  
 750 -FMEDSI-POFRIDAD---TGAVTTOAEALYED-QVSYTLAITARDNGIPKSDTYLE 802  
 1636 GIIGVGVGSFEIVADPLTLIGIRIARPLDYSPMKYTLITVYDNGSGSQGTITT 1695  
 803 ILVNDVNDNAPQPLDSDY-QGSVYEDVPFTSVLQISATDRDGLNGRFTYPOGGDDG 861  
 1696 IDVLVNDVAPTFNPALVYSANVLBSVDVGTSGTVSASDPEAGVFGQTFSIDPASDP 1755  
 862 GDFIVESTSGI-----VTRLRLDRENVAAQVLRVAVDQKMPARTPMEVTVTVLDV 914  
 1756 GLEFSEIISPIGAVNRADYKAKPLDRENVETHVLTLRATIDGVPPLTGYATLANTLDDV 1815  
 915 NDNPPVFEQDEPDVYENSPIGLAVARVATDPDEGTNAQIMYQIVBGNIPVFOQLDIF 974  
 1816 NDTPTFPARD-YRPTIKENILAEVQFYGISAKOBDP-TGCPPTLVLEVENT---DTDSF 1869  
 975 S-----GELTA-----LVDDYERPREYVLTQATSAFLYSKATVH---VLLDRDN 1019  
 1870 TWFVDVGGTTGNMNLISSKAISPDRETOATYSIPIRITDAPNGSGVNLIIYEIDBND 1929  
 1020 PVLGNFELFNNVYTNRSSSPFGAIGRVPAHDPDISLTYSPERGNELSLV----- 1074  
 1930 PHFGTTKELVYSFBN-----IPDSIGVGVDDKDTLEDKTY-----PELSDVYKQYR 1981  
 1075 INASTGEKLDSRALDNNRPLKAIMSVLSD-GVHS-VTAQCALRVTTITDEMLTHSTLR 1132  
 1982 VVEDTGQVITI---LAGTPAGTYEFVDRVADGVHADQISTVIVDVADIPREAVRSSGSPR 2038  
 1133 LEDMSPERLS-----PLGLGFIQAVATLATPPDHVVVNNQORDDAPGHLINLSL 1185  
 2039 FEGVTAEBELISVPSGGGQTKDLKGLIABEIIIPAKENVDIFSVINVPQP--TTVDVRY 2096

1186 SVGQPPGGGPPFLPSEDLQERLYANRSLTLAISAGRVLPEDDNICLR-PCENYMRGV 1244  
 2097 S-----AHGSPYTPPEOMGALANDRDLAOLAGVTIGMIKIDMLLESACES--ACT 2147  
 1245 SYLRPSSAPPLASSSVLFRPIHPVGLRCRCPGPGTGYCETEVLDLCYSRPGHGR 1304  
 2148 NVLBIIDPQVYINRPSAFVAVTSTVAKVC----- 2179  
 1305 SREGYVTLCDGTYGHECHVSASBGRCTPRVCNNGTCVNLVNGGPRCDGSSGPEKPY 1364  
 2180 ---GANT-----QQPGLCDPACLNGGTCCTETQGGGHTCCQCPG-YDGD 2220  
 1365 COVTRSPFASPIFRGLRORPHFTLALSPATKEDGLILNGRPF-----NEKDFVAL 1419  
 2221 CQQTREF-XGFAHFGTLQOCETHHSLEBITTRAEBVLXLYNGMIEVITGMDPDLMI 2279  
 1420 EYIQEQVQLTFSAGESTTVSPVPGVSDQOMHTVQLKYNNKPLLGQTLPOGSEQV 1479  
 2280 QLLGGQPVLEINTLSSGTLISLPATNLGDGKMRHLDV-YRN-----GKD 2323  
 1480 AVTVYDGC-DTQVA-LRGSVILANYSCAAQSTGGSKSLDLTGPIILGVPDLBSPF- 1536  
 2324 VEFMLDRCKDAVAVEASTSSSRQTDQCKITGOAPEDNKFLLVNTPLQSGI-DKSAGFTY 2382  
 1537 ---VRMGQFVGCNRNLQVDSRHIDMADFLIANGTVPCSPAKRQVCDSENTGNGGTCVNO 1592  
 2383 PDGITPAASTYGCNRNLEQDSTLYDLERPGAUSBAG--SKLYC--GECNN-GTCGCD 2437  
 1593 WDAPSCBCEPLGFGKSCAQAENANPOHFLGSSLVAMHGLSLPI---SQPWYLSMERTQA 1649  
 2438 FSTYCLCDPQYTBKCDQ---TPPAYDPATBSYVRQGLKPTIVIDSRGNVQVSPRTED 2495  
 1650 DGVLQAIT--RGSSTTLQLRBGHVMLSVGTGLQASLLEPRANDGMHHAQLAG 1707  
 2496 EGLMTWTNLNGLEFTTLEILNGYIR-SRWNLDDGHSMLDYQVAVNDGDMHWHLE-- 2551  
 1708 ASGGRHALISFD-YGQORAGNIGLPR-----LHGLHL--SNITVGGIIPGAPAGVAR 1756  
 2552 -----RFDYITVIRIDGGGGVQMESREBTFSTLEVDPSNLILGAFVHVHVDLSQ 2601  
 1757 GFRGLQGVRSVDTPREGVNSLDPSHGBS-----INVEGCSLPDPDCSNPCPANSYC 1808  
 2602 DFGCGMDPRINNYLAFHET-----ESDYAIPQPTASVTEGCSYDCAANPCVPFIC 2655  
 1809 SMDMSYSCSDPRY--YGDGCTNV--CDLAPCEHGSCTREKPARPHGYTCCEPNTYGP 1864  
 2656 KDVWRKTYICICKPPEBELGNTCVAILDDCDPNCINGICTDRDS--GFBCECPDGYRGD 2712  
 1865 YCETRI 1870  
 2713 ICDVAV 2718

RESULT 4  
 Q9WU10 PRELIMINARY; PRT; 4589 AA.  
 AC Q9WU10:  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Protocadherin.  
 GN FAT.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=99173791; PubMed=10072790; French Constant C.;  
 RA Ponsati M., Jacques T.S., Clant U., of the Drosophila fat tumour  
 RT "Expression of the rat homologue of the Drosophila fat tumour

RT suppressor gene."  
 RL Mech. Dev. 80:207-212(1999).  
 DR EMBL: AF100960; AAD20459.1; --  
 DR HSSP: P01132; 1EGF.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR002126; Cadherin.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR Pfam: PF00028; cadherin; 33.  
 DR Pfam: PF00008; EGF; 5.  
 DR Pfam: PF00054; Laminin\_G; 1.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 31.  
 DR SMART: SM00179; EGF CA; 1.  
 DR SMART: SM00282; Lamm; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00232; CADHERIN\_1; 17.  
 DR PROSITE: PS00268; CADHERIN\_2; 33.  
 DR PROSITE: PS00022; EGF 1; 4.  
 DR PROSITE: PS01186; EGF 2; 1.  
 DR PROSITE: PS01187; EGF CA; 1.  
 DR PROSITE: PS50025; LAM\_G\_DOMAIN; 1.  
 KW Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein.  
 SQ SEQUENCE 4589 AA; 505987 MW; 8CD0561BD58677 CRC64;

Query Match 9.8%; Score 1520; DB 11; Length 4589;  
 Best Local Similarity 26.6%; Pred. No. 4.5e-86;  
 Matches 478; Conservative 242; Mismatches 682; Indels 394; Gaps 57;

QY 179 POFQPSYQATVPENQAGTPVVALRAIDEGSAGRLTYM--DALFDSRNQPSLDP 236  
 DB 2702 PRSEPEPTYSISEDMPTGTEIDLIR-----HSGVLTLYTKGMPSNRBEFTVIDR 2756  
 QY 237 VTGAVTAAEILDNETSTHYF---RYTADHGMPPRSALATLTLYTDTNDHPVPEQG 292  
 DB 2757 QSRRLKLEKSLDHEHTKMYQFSILARCTLDYEV--VASIDVISQVKANDNSPVLESN 2813  
 QY 293 EYKESIRENLEVGEVLTARATDGDAPNANITLYRLEGSGSPSEVEFELDPPSGVIRTR 352  
 DB 2814 PTEAFIYENLPASRVYQVPSADSDSGVNGQVMTSLDOSODADIIISFALIMETGMITTL 2873  
 QY 353 GPVDRREVESYQLTVEASDQGRDPGPRSTAAVFLSYEDNDNAPOSEKRYVYVQVEDV 412  
 DB 2874 KEIDHEERASYQIKVAVSDHG-EKVQLSSTAIQVTVTVNDSPRTAIIYGTISEDD 2932  
 QY 413 TPQAPVLTAKTASDRDGS-NAVYHYSIMSGNAGQFLDAQTG--ALDVVSPIDYETKE 469  
 DB 2933 PPGGVLAIIISTTTADSEIRNOVSYPITGDDALGQFAVENMOMDWRVYVKKPLDRREQDS 2992  
 QY 470 YTLRVRAQDGRPLSVGLVTVQVLDINDNAPIFSTPFOATVLESVFLGYLVLAHYQ 529  
 DB 2993 YLLITVITDG---TSSKARVEYKVDANDNSVPCRTIYSDAIFEDALPGCLVMQVSA 3048  
 QY 530 IDADAGDNALRYLAGVGHDPFTINNGTGMISVAELDREEDYFSPVEARDHGTPTA 589  
 DB 3049 TDADIRENAETITTLFGSGAR-KPKLNPDTGELRTLLALDREQAVHLLVKAITDGG--G 3105  
 QY 590 LITASASVTVLVNDNPNFTPTQPEYTVRLANEDAAVTSVTVVTSANDRDA--HSVITYQI 647  
 DB 3106 RSCQATVTLTEVDNNDTPFTADPVAITVEFTEPEPTPLTRVOATDADAGLRKISYSL 3165  
 QY 648 TSGNTRRFSTITSOGGLVSLALPLDYKLEROYVLAVTASD-GTRQD---TAQIYVNV 703  
 DB 3166 LD-SADQFSSINQS-GIIQLEKHLDRLEIQAIVTTLTKADQGLPRKLTATGIVAVSVL 3222  
 QY 704 DANTRHVPQSSHYTVANNEDRPAGTTVVLISATDEDTGENARITY--FMEDSIPOFRID 761  
 DB 3223 DINDNPFVEFREYREGATVSEDIIVGTETVLQVYASRDIRANAEITVAIISGNHKGKSID 3282  
 QY 762 ADTCAVTTQALELDEVDQVSTLAIATARDNGIPQKSDTTYELILVNDVNDNAQPLFADSYQ 821

DB 3283 SKTGALFIITENDYESSHGVTYLTVEATDGTGTSLSDVATVININTDINDNSVFGSDYST 3342  
 QY 822 GSYVEDVPPFTSVLQISATDRDSGLNGRVFYTFQGGDDGDPFIYESTSGIYTRLRDR 881  
 DB 3343 TVSEDALEQPFVITIMDDAGPNSHILYIIBGNQ-SPTTIDPVAGEIKVKKPLDR 3401  
 QY 882 ENVAQVTLAAVAVDQKMPARTPMEVTVTLVNDNPNPFVEDEDFVEENSPIGLAVA 941  
 DB 3402 FTISGTYTLVQAADNGNPPRVVTTVINIDSVNDNAPLFSIDNYSVIIQENKPGFSTL 3461  
 QY 942 RVTAIDPEGTNA-QIMQIVEGNIPVFFQDIFSGELTALVDLDEBPEYVLVIAQ-- 998  
 DB 3462 KLVVTDKQSHNGPPPSFAIVSGDNDNEBVQHGVLTA-ATVRKAVDHYLLAKVAD 3520  
 QY 999 TSAPLVSRP-TVHLLDRNDPVLGNFELLFNNTVTRSSPFGCAIGRAPADPDS 1057  
 DB 3521 NKPQSLSTHIDIRATESHPALPLEI---FTASGEYEGGVIQKTHATDQVY 3576  
 QY 1058 DELTYSFERGNELSVILNASTGEKLSRALDNNRPLEAIMSVLSDGVASTAOCALRV 1117  
 DB 3577 DLTYSIDPRND-GLPVSSTGGKLIARKDIG---QYLVAVSTYDGKFTVADIIVHI 3632  
 QY 1118 TITTBMTLTHSITLLEDMSPERFLSPILGFIQAVATLATPDPHVVENVQDRTDAPG 1177  
 DB 3633 QQVQEMLNHTIATIRFANLTPEEFVGDWYRNQFARLNTIGIRKNDIQVSIQ--PSEBP 3690  
 QY 1178 GHILVNSVQCPQPGGPPFLPSEDLQERYLNRSLTALISARVLFPDNICLRBPC 1237  
 DB 3691 SH-LDVLFEVK-----SGTQISTKQLHKNSSVTDBEITIGVRLIDFQKLCGLDC 3744  
 QY 1238 -----ENYRCVSVLRPDSAPFLIASSVLPFRPIHPVGLRCR---CPGPGFG 1282  
 DB 3745 PKFCDEKVSVDENIMSTHSTARLSFTVRHRTAVCL-----CKDGTCPVHQG 3794  
 QY 1283 DYCEBTEVDLCYRPGPHGRCS--REGTYTLCEDGTGEGECVSAISGRCTPGVCNKG 1340  
 DB 3795 -----CEDNPPAGSECAVDAREKTCVCV-----PG----- 3821  
 QY 1341 GTCVNLVGGFCDDCPSGDFEKPQCVQVTTSPSPASFTFRGL--RQRFHTALSFAIK 1398  
 DB 3822 -----GGF-AKPGS-----SITTTGSSPVYKRLMBENRLEMLTNRIRY 3863  
 QY 1399 EEDGILLYNGREKHFVVALEIOVQLTFESAGESTTVSPFPVPGVS-----DQO 1451  
 DB 3864 SSHAVVMY---ARGDYSLIEHTGRQYKEDCSG-----PGIVSQSIQVNDQO 3911  
 QY 1452 WHTVQLKTYNKEFLQGTGLPGPSQKVAVTVVDGCDTGVALRFGSVLGNYSQCAQGTG 1511  
 DB 3912 WHAVSLV-----EGNYAKVLDEVHT-----ASGTAP 3939  
 QY 1512 GSKKSLDLTGPLLGGVPDLPSFPVRMQ-----FVGCNRNLQVDSRHID 1557  
 DB 3940 GALKTLANDNYIIFGG-----HLRQGTGRNTQVANGRGCK-----DSIYLN 3984  
 QY 1558 MADFIANN-----GTVPCC--PAKNVCDSDNTCHNGGTC-VNQMADFSCBCEPL 1602  
 DB 3985 GQELPLNNPRAVAHIEBVVDLAHGLLTATBD-CSSNPCQNGGVGNPSPGTGYCKCSA 4043  
 QY 1603 GFEGKSCAQBMANPQHFLGASSLVAMHGSLPISQWYLSIMRTQADGVLLQAITRGSS 1662  
 DB 4044 LHAQTYC----- 4050  
 QY 1663 TITLQLEBGMVLSYEGTGLQASISRLERGRANDGDMMHQAQLALGASGPGHAIISFDYG 1722  
 DB 4051 -----EVSVPNCSSN----- 4060  
 QY 1723 QQRABGNLAPRLGLHLNITVGGIPGAGVYAKRGKGLQGVXVSDTPEGVNSLDPSHG 1782  
 DB 4061 -----PCLYG-----GTCVNDNGGVCOCKRGLYTG----- 4085  
 QY 1783 ESINVEGGSLLPDPDSDNCPANSYCSNDMDSYSGSCPGYYGDNCC---TNYCDLNPCEH 1839



Db 4086 -----ORCOLSPYCDPCKNGTCTPDSLDGAVCCDSGFRGRCOSIDDECAGNFCRN 4139

QY 1840 QSVCTKRPSPAHG-YTCECPNLYGPYCETRIDOPCRGWWGHPCTGCPNCNDSKG 1894

Db 4140 GALTEN-----THSGYHCNCSOEYRGKHCEDATP-----NHVSPFMNIGLALG 4183

RESULT 5

Q8ABX7 PRELIMINARY; PRT; 4610 AA.

AC Q8ABX7; 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Procadherin.

GN FAT.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

RN (1)

RP SEQUENCE FROM N.A.

RA Down M.L., Ralston K., Smith S.I., Burns G.F., Boyd A.W., Power M.;

RT "Zebrafish Fat Protocadherin."

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF469608; AAO15696.1; -

SQ SEQUENCE 4610 AA; 506628 MW; 699B20BA4F6D3B93 CRC64;

Query Match 9.8%; Score 1519; DB 13; Length 4610;

Best Local Similarity 24.2%; Pred. No. 5.2e-86;

Matches 500; Conservative 248; Mismatches 674; Indels 640; Gaps 52;

QY 174 MNTABQPOPSPYQATVPENQAPGTPVSLRAIDPDEGAGRLTETMDALFDSRNQPS 233

Db 2381 NDWT-FVFSQNYEATVSELPARGHFTYQVQASDADISGKLF---SILSGNEQNTA 2436

QY 234 LDPVNGAVTAAELDETKSTHVFRTYADHGMPPRSALATLTLLVTDNHDPVFEQOE 293

Db 2437 MDNTALVTSNRRKPMESLIYNINVSVD-GVFRSAIYKAVNI--SANFHPPTNQVD 2493

QY 294 YKESLRENLEVGVEVLTVRATDGDAPNNAILYRLLEGSGSPSEPEIDPRSGVIRTS 353

Db 2494 YVVELLENSPVGLVLAQAQITDDSGTGRKTHIYINDIAKDFSISE---NGELFTLE 2549

QY 354 PVDRBEVESYQLTVEASDGRDPSPSTAAVFLSVEDDNDNAPQFSKKYVVQVEDVT 413

Db 2550 SLDR--NMLEKVIPISLAKOGGKVGFCVTAVIITLDINDNAPQFRADYRVNVAVDV 2607

QY 414 PGAPVLRVTRASDDKGSNAVHYSI-----NSGN----- 442

Db 2608 RGTTLIKIAASDMDGSSNMDITYSIETDADNVEGNPEIHQFSGVIYKESLIGLENQLYT 2667

QY 443 ----- 442

Db 2668 FLVRADGGNPTSSVVPVYVKILAEVVPKFTESHYRAIEDSLSGEIDVICTESD 2727

QY 443 -----ARGQ-----FYLDAQTGALDVVSPLDYETTKETTLRVAQD---GGRPP 484

Db 2728 QPVFSLVKGNTPESSNODBEVVDVPSGSLKDKDHBSTKMYQTLTDAQSYBSKAV 2787

QY 485 SNVSGLVTVQVLDINNAPIFVSTPQATYLVESVPGYLVLAQALDADAGNARLEYTL 544

Db 2788 SAAD--ISIQGVKNDRPLFESNPYEAVENTLPGTCVIQVKATDLDSTNGHYVYVL 2845

QY 545 AGVGHDP-----FTINNGTGMISVAELDEBEVDVFSFGVEARDHGTFA--LTAS 593

Db 2846 D-----PNEGRTDIYELFAINSESGWITTLKELDETRKTYVSLADRGKQVLAAS 2899

QY 594 ASVSVTLVDVNDNPPFTQPEYTVRLNEDAAVGTSVVTVSAVDRDA---HSVITYQITSG 650

Db 2900 TRVDVTVVDVNDNPPKFTAEIYKGTVSEBDDPGVIALISTTDLTEENNKQISYFTG 2959

QY 651 MTRNRPSTISGGGGLVSLALPLDYTLERQYLAVTSADGTRODTAQIVNVNTDANTHR 710

Db 2960 DPLGGFALIEHTQEMKVSVRKPLDREKONYLINTATGTGFAKAVEKVLADANSP 3019

QY 711 VPQSSHYTVANVEDRAGTTVLLISATDEDTGENNATIT--FNEDSLPQRIADTGAVTT 769

Db 3020 ICKESFYMESVPEDSPAGRLILQVSAITDINSMAISYELLDPGSEHSISDETEBLKT 3079

QY 770 QAEIDYDQVSYLAITATDNGIIPQSDPTVLEILVNDNDNAPQFLRDSYQSYVEDV 829

Db 3080 LPLDREBEHVHKMKRALDGG--GRPCAEVETIYEDVNDNPPQTTPTTYTTEENTE 3137

QY 830 PFTSVLQISATDRDSGLNGRVFTYFGSDGDDGFVSTSGIVRTLRLRENNVAQYVL 889

Db 3138 ITPVRLVYASDLDTGANRIMYSIL--DSADGVFSIEETGVVRLDRPLDRLOSLYTL 3195

QY 890 RAVAVDKGMPARTPE--VTVTVLVNDNPPPEQDEFPVPEBNSPIGLAAVATATD 948

Db 3196 RQATDRGSPRLHSQTLTSLVSLDINDNPPVFERREYATVAEDIPVGTQVLRVHAASR 3255

QY 949 DEGTNAQIMYQIVEGNIPEVQOLDIFSGELTALVDLDEDEPEYLVIOAT---SAPLVS 1005

Db 3256 DTEAGTEITVAILINGNERAFVVDPOTGVPVIEPLDIETAHFILTALDGGTSLSD 3315

QY 1006 RATHV----- 1011

Db 3316 LATVINILTVANDNSPFVQDIYSAVISEDAELGKTLVTMADADGFSNQVRFSLIDG 3375

QY 1012 -----RLDR-----NDNP 1020

Db 3376 NQSGPFTIDPVGEYKVALDLREKTSYTLTVLASDNGPARSSATTINVDSDINDNP 3435

QY 1021 PVLG-----NFEILFNN----- 1032

Db 3436 PIFSQANYSIIIOENQPIGASVQLTVSRDASHNGPPTFTSIKGNEDDFRITPGVL 3495

QY 1033 -----YVTRSS 1039

Db 3496 VSAATLSRQTOFYLQAVQVDSGRPLVSTAFVSVRLIEESITPSPILPLDIFITTAD 3555

QY 1040 SFGGAGRVPRADHDIDSLYSF-----EKGNELSVLNAASGELSLBALDNNRL 1094

Db 3556 EYSGGYLGIKATDQVDTLTYSLAPDSSSTSENSGLFSVSPAGKLVARGNLDG-- 3612

QY 1095 EAIMSVLSDGVSHTAQCALAVTITIDEMLTSHSTLRLEDSPERFLSPILGLFIOAVA 1154

Db 3613 QYVLNITVTDGFTAAARNIHVRQATAGALDNLSAVRSALAPBEFIDYWRNPLRALR 3672

QY 1155 ATLATPPDHVVVFNQRTDAPQGHILVNSLSVQPPGCGGPPPLPSDLQERLYNLS 1214

Db 3673 NIAGVRGQDVQLVSLQPAADASGD--LEVLTALER-----SGSPPOQEVLYRKLNASVG 3725

QY 1215 LTLTAISORVLPFPDNICLRBPCENYMCVSLARDSAPFLASSVLRPRLHPGGLRC 1274

Db 3726 VLEMTGVAIYAVVKKLCAGLDLP--LSFQETIATSTGA--VSAYSTARISFTPRMRTA 3783

QY 1275 RCPPEFTGDYCELEVLDVLSRPPCGPHGRC--RSRREGYTCLCDGYTGEHCVSARSGR 1332

Db 3784 KCP--CBAGBSGILSKLREGSPCPGMCVCEBPTHSKRTCLPE---GQDECSBSRGSLT 3638

QY 1333 TRGVCKNGGTCAVLVGGFKCDPBGDFEKPVCQVTTSSFPNASHSTTFPGLRQRFHTLA 1392

Db 3839 FSG--NGARVRLMENR-----EEMKLSRLRTFFSTHATVHVQORHLOH--- 3883

QY 1393 LSFATKERDGLLVNGRFRNEKADPVALFVIGQVQLTFSAGESTTVSPFVGVSQDGM 1452

Db 3884 -----TEHDDGPLQYK-----FDCCSGPGVSVSH--STQVSDGM 3916

QY 1453 HTVQLKYNNKPLGLQGTGLPQGSSEQKVAVTVVDCDGTVALRFGSVLNTS-----CA 1505

Db 3917 HTVSL-----VNGVYATLVLDQVYA 3937

QY 1506 AQTGGGSKSLDLTGPLLLGG--VPDLPSFPVVRKQVGVGCMRNTQVDSRHIDMADFLAN 1564

Db 3938 ASGTAGTTLRTLTMLDTSMTFGGAVRPAASGSLVMSGLRCLGIVNGRELPLSGVRGA 3997  
QY 1565 NGTV-----PSCPAKQV--CDSNTCHNGTGC-VNOMDAFSCBCEPLGFGKSCAQEMA 1614  
Db 3998 HAVLEELVDVAPRCSLAPRIQSTSNPCYNGTGTCSALPBGVYCKTAAFMGTHC----- 4052  
QY 1615 NPQHLGSSLVAMHGLSLPISQPMYLSLMEFRTRQADGVLLQALTRGRSTTTLQLRGGHVM 1674  
Db 4053 ----- 4052  
QY 1675 LSVETGGLASSLRLEPRGANDGWHHAQLALGASGPGHALLSPYGOQRAAGNIPRL 1734  
Db 4053 -----EVTISPCASN-----PCL 4065  
QY 1735 HGLHLNITVGGIPGAGVARGP-----RGCLQGVVSDTPREGNSLDPHGESINVEQG 1790  
Db 4066 Y-----GGTCIRGG--DFYCCQKQYSG-----OR 4089  
QY 1791 CSLDPDCSNPCPANSYCSNDMSYSQSCDPGYGDNCC--TNVCDLAPCEHQSVCTRRP 1847  
Db 4090 COLGPFCTNPNCKNSKCIDSLDGPVCEPFGQERCLSDVDECLNFCNGHCQNTY 4149  
QY 1848 SAPHGTCCEPPVYLGPRYCETR 1869  
Db 4150 GS---PSCNCSNGYSGQHCELR 4168

RESULT 6  
061230 PRELIMINARY; PRT; 2809 AA.  
ID 061230  
AC 061230;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE G-cadherin.  
OS Lytechinus variegatus (Sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Echinozoa; Echinodermata; Echinacea; Temnopneustidae;  
OC Lytechinus.  
OX NCBI\_TaxID=7654;  
OC RN  
RP SOURCE FROM N.A.  
RX MEDLINE=98104238; PubMed=9441671;  
RA Miller J.R., McClellan D.R.;  
RT "Characterization of the role of cadherin in regulating cell adhesion  
during sea urchin development.";  
RL Dev. Biol. 192:323-339 (1997).  
CC -FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR HSPB; U34823; AAC06341.1; -  
DR HSPB; P01133; IEGP  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR002126; Cadherin.  
DR InterPro: IPR000233; Cadherin\_C\_term.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR006209; EGF\_like.  
DR InterPro: IPR006210; IEGF.  
DR InterPro: IPR001791; Laminin\_G.  
DR Pfam: PF000028; cadherin; 16.  
DR Pfam: PF01049; Cadherin\_C\_term; 1.  
DR Pfam: PF000045; Laminin\_G; 2.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 17.  
DR SMART; SM00181; EGF; 3.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00032; CADHERIN\_1; 2.  
DR PROSITE; PS50266; CADHERIN\_2; 14.

DR PROSITE; PS00022; EGF\_1; 3.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 2.  
KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
KW Glycoprotein  
SQ SEQUENCE 2809 AA; 303041 MW; 60191A3B643BA0F3 CRC64;  
Query Match 9.7%; Score 1508.5; DB 5; Length 2809;  
Best Local Similarity 29.0%; Pred. No. 1e-85;  
Matches 538; Conservative 252; Mismatches 757; Indels 311; Gaps 87;  
QY 147 SCKLAQAGLRAGE-----RSPRESLGRKRRKRVNTAPQGPSPYQATVPEHNPAGRP 199  
Db 938 SCIVTASGIRSRVDVDFVRSPPDT-----TYGVPTFSNTPYQVTTINENTTIG-D 988  
QY 200 VASLRALIPDEGEARLEFYTMALPDSR-----SNQFSL-----DPVATVTAELDET 251  
Db 988 VLLVSV-----GAQGVGPDIVGSPSTMSDGTFTLTRDDNLQSDTLVEKLDRE 1042  
QY 252 KSTHVFRTVADHGMKPRRSALAT-LTILVTDTNDHPVPEQGEYKSLRENLEGYE--V 308  
Db 1043 VQEVLELVTAQVGVGQEFIAQTPTVTITEDPNDVAPKPAVTFYGNIAENQPVNSDPI 1102  
QY 309 LTVRATDDAPRNAILTRLLEGSGSPSEVEPEIDPRGVTIRGPVUREEVE--SYQLT 366  
Db 1103 ITLQATDDVTPPE---YSQISYSFVNABSPESIDEDGELFALRVFREALEDSIALT 1158  
QY 367 VEASDQGRDPGPRSTTAAVFLSVEDNDNAPQSEKRYVYQVREBVTGAPVLRVTAADR 426  
Db 1159 VRASD-----GVNQLATVFTITVDENDEBPN-GTSSFDVLEVGNGYDGTGTATDD 1212  
QY 427 DKGSNAVVHYSIMSGNARQFYLDAQALDVSPDLYETTKRYTLRVRAQDGRPPLSN 486  
Db 1213 D--ISEVLEVFISGNEGGAFTVDAEETIRAGVLDYEARTSYELQSVNDG-----KN 1265  
QY 487 V-SGLVYQVVDINDNAPLIPSTPRQATVLE--SVPLGVLYHQAQI--DADAGNARLE 541  
Db 1266 VATTVTNTVNVNVAAPQPSAASVIEBDSNLPRILISVAATQGDADAVDA--V 1324  
QY 542 YRLAGVGHDFPTTNGTGMISVAELDREVDYFSPGVEARDHGTPLATASASVTVL 601  
Db 1325 YGLVGTGAGTIFTDISQGNITLTQALDRBHPYTNLAAVATDNGNGLTSVVDVITIE 1384  
QY 602 DVNDNPPFTQPEYTVRLNEDAVGTSVTVYSAVDRDAHSVTTQITGNTNRPSITSQ 661  
Db 1385 DINDNAPYPPQGEYVGSVEERKPPPTPVAAVABPDADLMYFPPSP--DNNINSQ 1442  
QY 662 SGGGLVSLALPLDYKL-BROYVLAVTASDGTRODPAQIVNNTDANTHRPVQSSHY-TV 719  
Db 1443 T--GQITTAQDFRTPSEYEIEVQATDGVNTASTVATISIDVDNKPSPSEBVPYDPA 1500  
QY 720 NVNERDPAGTIVVLLISATDEDTGENARITVFMEDSIPQ--FRIDADT-----GAVTTOAEL 773  
Db 1501 SVLETPETGTTITVQALDIPDVPDPRDAVFSINSDDPELPRIVADAATLQGLIQNKEL 1560  
QY 774 DYEDQVS--YTLATARDNGIPQKS--DTYTLAILVNDVNDNAPQFLRDSYQSVYEDV 828  
Db 1561 DEETLATEPFLTVAVTOSQGTSGRPEIATVKKIIVNANDLAVFPQDVPDVGAVSBA 1620  
QY 828 PFTSV--LQISATPRD--SGLNGRVYTPQGGDGGDFVBSSTGIVTRLRLADREN 883  
Db 1621 TNSQVGFPTSATIDGDEFGATLN---YIIDPTTPDQGFAINENQOLI-VASPLDRET 1675  
QY 884 VAQVYLRVAVDKGMPARTMEVTVYLVAVNDNPPVEQGEFVFEENSPI--GLAVA 941  
Db 1676 VASYELKVVAVDNGPMPSGTVAVTVDVDTDPHQAQD-XNPSVEE-GRIBANVEV 1733  
QY 942 RVTATDPPE-GTNAQIMQIVEGNIPVFOQDIF-----SGEL---TALVDLYE 987  
Db 1734 SYEAVDDDDPSPGPPFLYNVA-----POPNDWTFPDIIRGLGTSGSIRVSTTGLEIRE 1789  
QY 988 DRP--EYVLVLDATSAP--LVSRATVHVELDRNPNPVLGNFELFNNAVYVTRSSSPFG 1043

```

Db 1790 THEPYDIVFLIAEVTPEALTGOTLTIMISDVNDNPHVATKOLLVSEYEN----1PT 1845
Qy 1044 GAIGRVPAHDPDISLTYTSFERKNEI-SLVILNASTGELKSRALDNRRPLEAI-MSVL 1101
Db 1846 TEVGKTKVEDEPDLEKTY--EAVGILPDFQOLDSTGDTMA---EGTACVTEMDR 1899
Qy 1102 VSDG--VHSTVACALRVITITDEMTHSTTLLEDMSPERFLSP-----1LAGFL 1149
Db 1900 VSDGTYESVSVIVITVKDIPREAVSSGSIRFSGTMBELITPSEGSNDRLKXVL 1959
Qy 1150 IQAVATATAPPHVNVV--QRTDAP-GGHILANSLVSCQPPGCGPPPLSEDL 1205
Db 1960 ANAVGAQLAN-PDIFSVLVNNGERTVDIRYAAH-----GSYYPADOL 2002
Qy 1206 QERLYNRSILTAISAOR-----VLPPDNICLAE-PCENYMRCSVYRPPSSAP 1254
Db 2003 D-----LAALSTSNIEBELGISIAQIPVD--LCVENAVCES--SCTNVLVVDPTPT 2049
Qy 1255 FLASSSVLFRPIHPVGLRCRCPGFTGDCETEDVLCYSRPGPHGRCSRREGYTCLC 1314
Db 2050 VVDSGTASLVAITSV--LEAQC-----ICGARITVAF----- 2078
Qy 1315 RGATYGEHEVSNARSRCPTGVCNKGCTCVNLLVGGFKDCSGDPERKYCVVTRSPRA 1374
Db 2079 -----GNCSDSPCLNGCTCEDVHGCTCTCPY-LFDGPNCOQTKRSF-B 2121
Qy 1375 HSFTFRGRORPHFTALSPATKRDGLLVNGRF-----NEKHDFVALVTOGVOULT 1429
Db 2122 NGVASTIRROCEBESLSEIFTEVSSGTLNNGPFTPTGDDPIMILELGGAKLT 2181
Qy 1430 FSAGESTTT-----VSPFVPGSVSDGQWHTVOLKTYNRPPLAQTLPOGPSBOKAVAVT 1483
Db 2182 INIGSTSDTDMVLLEAPTDPTDQNLNDNEMHRIDV-YRN-----GRVEMT 2225
Qy 1484 VDGCGTGV---ALRFSSVLGNVSCAAQCGQSGK--KSLDGLPPLGGV-PDLPSFPV 1537
Db 2226 VDRC-NGVFPFAETSSSTLDTSSCRVNGTTPGENMLRFLNVHTPLFLGMSADYDVTVP- 2283
Qy 1538 RMRQFVGCNRNLQVDSRHIDMADFIANNGTVPSCPAKKNVC-DSN---TCHNGTGVNQM 1593
Db 2284 ---SGPDGCIKNLVSDGFLYDLGTFPGTSSSEKACPTDQCTDDNMPVCNN-GTCEADI 2340
Qy 1594 DAPSCBCLPAGFGKSCAQEMANPQHFGLSGLVAMHGLSLPI---SQPWTLSLMFTRQAD 1650
Db 2341 DSFCICFPFGNGLTCDVEL-TPYDPAISYITVELDSSLVDADRASSNYQLMWTROEN 2399
Qy 1651 GTLLQALITGRS--TITLQREGHVLSTVEGTGLQASSLRLRPGRANDGMH-----H 1701
Db 2400 G-LIMSISANTYETIRMEMVQELKADWH-LGDKFVSVTWVNFSLINDGAMHAINFDRYD 2457
Qy 1702 AQLALGASGPGHALISPDYGOORABGNLGPLHGLHL-SNITVGGIPGPAGVARGFRG 1760
Db 2458 SVYTIKIDGGGKVEIQ-----NRESQYSGLDVENSLLVGAFTVDVTTDDDFG 2507
Qy 1761 CLQGVAVSDT---PEGVNSLDSHGESINVEQCSLPDPCDSNPSPAN-SYCSNMDSYS 1816
Db 2508 CANDRIRINFLMGEGINDYAVA-TKSAGVTEGCP-SDVCDSDFCGSLVCTDWMRFYE 2565
Qy 1817 GSCDDGY-----GDNCTNVCL--NPGHOSVCTRKBSAPHGYTCECPNTLGYCE 1867
Db 2566 CLCPGQEEVEDDPITCAIIDCVNPPCANGTCAVE--GDPGYTCDPCSGYGYDRCE 2621

```

## RESULT 7

```

090X3 PRELIMINARY; PRT; 4587 AA.
AC 090X3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 13, Last annotation update)
DE Mouse fat 1 cadherin (Fragment).
GN PATH OR MPAT1.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=embryo;
RX MEDLINE=20203461; PubMed=10741417;
RA Cox B.T.M., Hadjantonakis A.K., Collins J., Magee A.I.;
RT "Cloning and expression throughout mouse development of mfact a
RT homologue of the Drosophila tumour suppressor gene fat.";
RL Dev. Dyn. 217:233-240(2000).
DR EMBL; AJ250768; CAB65271.1; -.
DR HSSP; P08709; 1BP9.
DR MGD; MGI:109168; Fatch.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00028; cadherin; 33.
DR Pfam; PF00028; EGF; 4.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 31.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00282; LamG; 1.
DR PROSITE; PS00232; CADHERIN_1; 17.
DR PROSITE; PS00268; CADHERIN_2; 33.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS00025; LAM_G_DOMAIN; 1.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
FT NON TER 4587 4587
SQ SEQUENCE 4587 AA; 506036 MW; 4D3F23B05127CB4 CRC64;

Query Match 9.6%; Score 1496; DB 11; Length 4587;
Best Local Similarity 23.5%; Pred. No. 1.5e-84;
Matches 533; Conservative 267; Mismatches 662; Indels 810; Gaps 68;

Qy 179 POPOPSTQATVPENQAPGTPVASIRAIIDPEGEAGRLLEYTMALFDSRNOFSLDPT 238
Db 2283 PVFQPSISTTLSEASVGTIVLOVRATDSSEPNRGISTYLIG-NHSSKHDFHIDSNT 2341
Qy 239 GAVTTAEELRETKSTHVFRTAODHGMPPASALATLITVTDNDHDVPVEQOEYKESL 298
Db 2342 GLISLVRALDYEGQQRHIFRAVDGKPLASDVVTAVNTDLDNDNPLEEQVYEAR 2401
Qy 299 RENEVGYEVTLTPATGADAPPNANILYRLISGSGSBESEVFETIDPPSGVT----- 349
Db 2402 SEHAHGHFWCVRACDADSDLDLKEYSIL---SGMDHKSFIIDREGTITLSNLRHT 2458
Qy 350 -----RTRGPV-----DREVE-----SYQLTVEASD 371
Db 2459 LKPYISLVNSVSDGVFRSSARVNVTVMGNLHSPVPHONEYELAENAPLHTLVQVKA 2518
Qy 372 QGRDPG-----PSTTAAVFLSYE----- 390
Db 2519 SDRDSGIYSHVTHIIVNDPAKDRFYVNDRQIPTLLEKLDRETPAKYISIRLMAKDA 2578
Qy 391 -----DNDNAPQPSSEKRYVQVREDVTGAPVLRVLTASDDKSGSNVAVHYSI- 438
Db 2579 VAFCTVNVILTDNDNAPQFSTKYEVNIGSSAKAGTSVVKFASDADBSNADVTYAI 2638
Qy 439 ----- 438
Db 2639 ADSESVKENLEINGLTGLITTKSLIGLENFFTFPRAVDSGSPRESVVPYIKILPP 2658
Qy 439 -----MSGNA-----RGQFY-LDAQ 452
Db 2699 EVQLRFRSPPFYTYTISEDPIGTETIDILRVHGAVLITLVKGNTPESNRDEFVDRQ 2758
Qy 453 TGLADVSPIDYETTKETYLRAVQDGRPLSL--VSGI-VTVQVLTDINDNAPIVST 508
Db 2759 NGRLEKSLDHTTKYQPSILA---RCLTDYEVVASIDVSIQVKANDNSPVLESS 2814

```

QY 509 PFOATVLESVPLGVLVLAHVAIDADAGDNARLETRL-----AGVGHDPFTTNGTGMIS 563  
 Db 2815 PYBAFVLENLPQSGSRVQIRASDIDSGANGVWYSLDQSOADADIES--FAINMEKMIT 2872  
 QY 564 VAALBREEDVDFSPGEARDHGRPA-LTASASVSVLVLDVNNNPPFTQPEYVLANED 622  
 Db 2873 TKELBHEERASQIKVVASDHEKVKQSLAIVGVTVTDVNSPPFTMEIKGTVESD 2932  
 QY 623 AAVGTSVVTVSAVDRDAHVS---ITYQITSGNTRNRPSTISOSGGGLVSLALPLDYLER 679  
 Db 2933 DPGGVTAISTTDADTEINRQVSYFTTGDDALGQAVENVOGDMRVYKFKRLDREQD 2992  
 QY 680 QYVLAATASDGTRODTAQIVNTDANTHRPFGSSHITVNVNEDRAGTTVLISATDE 739  
 Db 2993 SYLLTVATDGTSSSKARVEVKVLDANDNSPVCEKTSYSTIIPEDALPGCLVWQVSATDA 3052  
 QY 740 DTENARITYFMEBS-IPFRIDADGAVTTOAELDEDOVSYTLAITARDNGIPQKSDT 798  
 Db 3053 DISNMEITTLTGSGSKAKTKLPDICEKTLALDDEQAVNVLKATDGG--GRSCQ 3110  
 QY 799 TYLEILVNDVNDNAPQFLRDSYOGSVYEDVPFTSVLQISATDRSGLNGRVFTYRQGD 858  
 Db 3111 AAILVLEEDVNDNAPERTAPYTITVEENTEPTLFRVQATDADGLANKISYSLV--E 3168  
 QY 859 DGGDPFVSTSGIVTRLRLDRNNAQYVLRAYNDKMPARTPM-EYTVVLDVNDN 917  
 Db 3169 SAQGPSINERSGIIQLEKHLDELQAVYTLTKAVDQGLPRRLTATGVVAVSLINDN 3228  
 QY 918 PVPFODEPVPVPEENSPIGLAVARVATPDDEGTNAQIWOIVGEINPEVFDLDPSCG 977  
 Db 3229 PPFETREYKASVSEDVIGTEVLQVYASRDIBANAEITYAISGHEKFSIDSKTGA 3288  
 QY 978 LVALVDLDYEDREPEYVLQAT--SAPLVSRATVHRLDRNDNPVLG----- 1024  
 Db 3289 IFTIESLDYESSHYULTVEATDGTPLSDVATVININVTIDINSPVFQDTYTTVSE 3348  
 QY 1025 ----- 1024  
 Db 3349 DAALBOFTITMADADAGPSNSHIYSILEGNOGSPFTIDPVGEVKYTKPLDRETTISG 3408  
 QY 1025 -----NPEI-----LF--NNY-----VTN 1036  
 Db 3409 TLTVQADNNGNPRVNTTNNIDVSDVNDNAPLFSRNYSVIIQENKPVGFSVLKLVVD 3468  
 QY 1037 RSSS----- 1040  
 Db 3469 KDSHNGRPFFFTIVSGNDENAEVNGHVLTPATIKRKYKHFFLVAVKADSGKQLS 3528  
 QY 1041 -----PPGALGVNPAHNDPDISLTYSPERGN 1068  
 Db 3529 SMTHIDIRVIESIHPPAILPLEIFITAFGEBSYSGVIGKIHATDQVDTLMTYSLDPHM 3588  
 QY 1069 ELISVLINASTGELKLSRALDNNRPLEAIVSVDSVHSVTACALRVITITDEMILHS 1128  
 Db 3589 D-GIFSVSSTGKLIARHKDIDG---QYLLNVSTDSKFTTADIVTHIQOVQEMLNHT 3644  
 QY 1129 ITTLELMSPEERLISPLGLPIQAVATLATPPDHVVVAVVAVQDTDAPGGHIIANVSLISV 1188  
 Db 3645 VAIRFAMLTPEERFVGDTWRNFORALRNILGVKKNDIQIVSLQ--PSEFHSN-LDVLLFV 3701  
 QY 1189 QPFGGGGPPFLPSEDIQERLYNRSU--LTAISQAVLFPFDNICLRBSC----- 1237  
 Db 3702 R-----SGGTTFTSKQLAHK--INSSVTDBEILIGVILLEVQKLCGLDCEPMKFCDEKY 3754  
 QY 1238 -----ENYVRCSVLAIFDSSAPFIASSSVLFRPIHVGIGRCR--CPFGTGDYCETEVDL 1291  
 Db 3755 SVDENVMVSTHSTARLSFVTPRHRTAVCL-----CKDGCPRVNHG----- 3795  
 QY 1292 CYSPFCPPHRCMS--REGGYTCLCRDGYGHEGCVSASGRCTPCYCKNGGTCVALLVG 1349  
 Db 3796 CBNPFCPAGSECVADPREBKYSVCV-----PG-----G 3823  
 QY 1350 GFKCDSPSGDFEKYCVVTTSPFASHTFTRGL--RQRFHPTLASFATKERDGLLY- 1406

Db 3824 GFG-KCPGS-----SSTFGNSFVKRYKFLLENENRLEMKLSMRLTRYSIAHAVMYA 3873  
 QY 1407 NGRFNEKHDFAVLEVIQEQVOLTFSAGSSTTVSPFVPGVSDGQOMHTYQYKPLLG 1466  
 Db 3874 KGTDYISIPGISVQSIQ-----VNDGQMHAVISEV----- 3903  
 QY 1467 QTLPGQBEQKAVAVTVVDCDPTGVALRFGSVLGNYSCAAOGTQGSKSLDLTPLLLG 1526  
 Db 3904 -----EGNYALVLVDEVHT-----ASGTALGALKTLINDVYVFG 3938  
 QY 1527 GVPDLPESPFVPMRQ-----FVGCERNLOYDSR-----HI----- 1556  
 Db 3939 G-----HLRQGTGKGAGVAGVAGFRGCDMSIYLANGQELPLNKPRAVAHIEWV 3988  
 QY 1557 -----DMADFIANNQTV--PG-----CPAKNVCDGNTCHN 1585  
 Db 3989 DLSHGCLLTATBDCSSSPQONGVGNPSPGTGYCKNALYVGTCEVSVNPPCSSNPCLS 4048  
 QY 1586 -----GGTCVNMQDAFSCPCPLGFGK 1607  
 Db 4049 GGTFTICDSGVCSGLVILCSDQLCSFQDDPCQIGTGFPSDLGAVCHDGGFRGE 4108  
 QY 1608 -----SCAQEN-----ANQHPUGSSLVAMHGLS 1631  
 Db 4109 RFQIDIDECAGNPCRNKALCEITLTSFYHCNCSQFGRKRCBDASPNHY----- 4156  
 QY 1632 LPIQPMWYLSLIFRTRQADGVLLQAITGRGRTITQLREGHVMSVEGTGLQASSLRLEP 1691  
 Db 4157 --VSTFPMNIGL-----AEGIGIYFVINGIYLVNVVPLCKRMIS-----RKKK 4197  
 QY 1692 GRANDQWHAQALGASGGPCHALISPDYQGOARBNIG--PRLGHLSNITVGIP 1748  
 Db 4198 ROAEPEB-----KRLGPTTALQRPYFSPSKINKAIYFDIPAOVVRPISYTPSIP 4248  
 QY 1749 GPAGVAAAGFRCLQGVAVSDPBEVNSLDP-----SHGESINVEGCSL-----PDQDS 1799  
 Db 4249 SDSRNNDL--RNSPFGSALPHEBP--FTFNPSMAGHKAAYV--CSAANPLPPLPPSN 4302  
 QY 1800 NPPCANYSYCSNDMDSYSCSDPGYGGDNCTVNCIDLNPECHOSVCTRKSPAPH 1851  
 Db 4303 SPDSDSIQKPSWDF-----DYDAKVVDLDPCLSKKPLBKSGQPY 4343  
  
 RESULT 8  
 Q8R508  
 ID Q8R508 PRELIMINARY; PRT; 4555 AA.  
 AC Q8R508;  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE PAT3.  
 GN PAT3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=21670969; PubMed=11811999;  
 RA Mitsui K., Nakajima D., Ohara O., Nakayama M.;  
 RT "Mammalian Fat3: a large protein that contains multiple cadherin and  
 RT EGF-like motifs".  
 RL Biochem. Biophys. Res. Commun. 290:1260-1266(2002).  
 DR EMBL; AB076401; BAB86869.1; -.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF 2.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR001791; Laminin G.  
 DR Pfam; PF00028; cadherin; 33.

DR Pfam: PF00008; EGF: 3.  
 DR SMART: PF00054; laminin G: 1.  
 DR SMART: SM00112; CA: 32.  
 DR SMART: SM00179; EGF CA: 1.  
 DR SMART: SM00282; lamG: 1.  
 DR PROSITE: PS00010; ASX HYDROXYL: 1.  
 DR PROSITE: PS00032; CADHERIN\_1: 19.  
 DR PROSITE: PS00268; CADHERIN\_2: 30.  
 DR PROSITE: PS00022; EGF\_1; 3.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF CA: 1.  
 DR PROSITE: PS00025; LAM G DOMAIN: 1.  
 DR Calcium-binding; Cell adhesion; EGF-like domain;  
 Glycoprotein.  
 KW  
 SQ SEQUENCE 4555 AA; 502077 MW; B47CB010638BA73B CRC64;

Query Match 9.1%; Score 1413; DB 11; Length 4555;  
 Best local Similarity 26.1%; Pred. No. 2.7e-79;  
 Matches 520; Conservative 264; Mismatches 807; Indels 400; Gaps 72;

179 PGPQPSYQATYVENOPAGTPVASLRALDPDEGRAGRELYTMDALPDSRNOFPSLDPVT 238  
 2705 PFTQSQYSEFTLAEDTSIGSTIDTLRL--PQSVRFSTVNGERENKKNFIIIEQST 2761  
 239 GAVTAEELDRKSTHVRVTRQDHGMRRSLATLT---LVTDTNDDHDPVEEQEY 294  
 2762 GAIKLDKRLDHEVSPAHEKVAAT--IPLDKVDIVFTVDVVKVLDLNDKPEFTSSY 2818  
 295 KESLRLENLEVEYELVTRATDGDAPNPANILYRLLEGSG-GSPSEVFEIIPDSGVRRKG 353  
 2819 ETLTMEGMVGTKLQAVRALDIDMGANGVYTSLSHSDSHLEKMEAFNIDSNITKSTLK 2878  
 354 PVDREVESEYQLTVEASDQGRDPRSTAAVFLSYEDNDNAPQSEKRYVVOYREDVT 413  
 2879 DIDHETDPTFSEFVVASDLG-EAFSLSMALSVKATDINDAPYFAHEVYNGVMSBP 2937  
 414 PGAPVLRVTRASDRDKS-NAVVHYISMGARGQFILD--AQTGALDIVSPIDYETTKEX 470  
 2938 PEEVAVASTLDKDTSNIRQVSYHITGPNPRGFRALGMVQSEMKVYVRPLDRESDIY 2997  
 471 TLRVRAQDGRPPLSNVGLVTVQVLDINDNAPIFSTPFOATVLESVPLGLVLAHVAI 530  
 2998 FLNITRASDS---LFTQAMVEVTVSDVANDSPVCQVAYSASLEPDISNKIILKVSXK 3053  
 531 DADAGDNARLEYRLAGVGHDPFTTNGTGMISVAELDREEDPYSPGEVARHDTPTAL 590  
 3054 DADIGSNGDIRXSLVSGNS-DEFLDPESGELKTALLDRERVPYNNILARATDGS--GR 3110  
 591 TASAASVTVLVNDNNPTFTQPEYTVRLNEDNAVGTSVTVSAVDRDA--HSVITYQIT 648  
 3111 FCSSTVLLLEEDVNDNPPVFSNHYTACVYENTATKALLTRVOADVDPVGINRKVYSI- 3169  
 649 SGNTNRFSITQSOGGLVSLALPLDYKLEQVYLAVTASD---GTRQDTAQIYVNTVD 704  
 3170 EBSAGVRSIDSS--GVIVLEQPLDREQSSYINISVRATDQSPGQSSLSLTSVITVLD 3227  
 705 ANTHRPVQSSHYTVANNEDRPAGTTVVLISATDEPTGENARITYEMEDSIPO--FRIDA 762  
 3228 INDNPPVFERRDYLVTVPEDTSLGTVQVLSVFATSDIGTNAITLILINSNGQKFRINP 3287  
 763 DTGATVQAELEVEQVSTTALITARDNGIPQKSDTTIYELIYVNDVNDNAPFLDSTQ 822  
 3288 KTGGSIVLEALDYEMCKRYLLVEAKDGTPLSTAAVTSIDLTVNDNPPRFSQDVSA 3347  
 823 SYVEDVPFTSVLQISATDRDGLNGRVFTYFQGGDGDGDFIVESTGIVTLRLRLRE 882  
 3348 VASEDALBEDSDYILIAEDVDSKPNQIRFSIVG--DRNEPAVDPIGLGVYKGLDRE 3406  
 883 NVAQVLRAYAVDKMPAPRTMEVTVTVLVNDNPPVFEODEFVVEENSPIGLAVAR 942  
 3407 RVSGVSLILQAVDSGIFPAMSTTTVVIDISVDNDSPVFTPANTYAVIGENPVGSTLIQ 3466  
 943 VATATPDEGTNA-QIMYQVBEGNIPVPLDIFSGELTALVDLDVDEPREYVLVQATISA 1001

DB 3467 LVVTRDRSEPHNGPPPSFISLGNDEDEEFLMD-SHGILSAVAFRMESEYLLCIQAKDS 3525  
 1002 ---PLVSRATVAVRLDNDNDNPVLGNFELLFNNTVTRSSFPFGAIGRVAPHDPISD 1058  
 3526 GKPQGVSTHYIRVRVIEESTHKEPTAIPLFI---PIVWEDFPFGVGIGKHATQDMVD 3581  
 1059 SILTSPERGENELSLVLANSTGELASRALDNNRPLEAMSVLVDGVSVTAQCALRYT 1118  
 3582 VLTFL-LKSBQSLPLKVNASHGKIIALDGLDSGK---YLVANVSVDGRFQVIDVAVNE 3637  
 1119 IITDEMTHSTTLRLIEDNSPERFISPLIGLPIQAVA---ATLATPDHVVENVQSDT 1173  
 3638 QLVHEMLQNTVTVIRFENVSPEDF---YGLMHGFRILIRNVALVLOKODSLRISIQ--- 3690  
 1174 DAPGHIANVLSVGPPPGGPPPLPSBDLQERLYNRL-LTATAGVLPFPDDNIC 1232  
 3691 PVVGNTQMDMLPAVEMHSSEYTKPAYTLQKLSMARHLENVHTALIERKCSGLD---C 3747  
 1233 LREPCENTMRGVSVLRFPSAPFLASS-VLFRPIHPVGLRCRCPFGTGDYCEYVDL 1291  
 3748 QEQHCE-----QGLSDSHLMTYSTARISFVCPRFYRNRCTC---NGGVCESNDP 3797  
 1232 CYSRPCGPHGCRSRREGYTCLCRDGTGHECEVSARSGRCTPGVCNKGTCVNLVGF 1351  
 3798 CVERKCPEDMOC-----VGY---EASRRP-----F 3819  
 1352 KDCPSGDPEKPYCO-VTTRSPFASPTTR---GLQRHFHTLALSPATBERDGLALYN 1407  
 3820 LQCCPGLGGE--CSGHTSLSFAGNSYIKYLSSENSRE-DFKALRLRITQSNQIMY- 3875  
 1408 GRFNEKDFVALLEVIOGV--QLTFSAGESTTVVSPFPGGVSDGOMHTVOLK----- 1458  
 3876 ---TRANCMILKIVEGLMWQLDQSGPGLIGIS---SRVNDGSMHSVLELNRTS 3929  
 1459 ---YNNK---PLIGCT----- 1468  
 3930 LSLDSYVERRRAPLYFQTLSTDSALFPGALVQADNIRSLTDTVTQVLGGFQGLDSVV 3989  
 1469 ---GLP-----QGPSEOKV-AVVVVDGDDGVALARFSV---LGANSCA-- 1506  
 3990 LNHHELPQNRSSFAEVGTELTGCVLPDQCSNPCLHGSCSGLPBGSTQSCLS 4049  
 1507 ---QGTGGSKSLDLTPGLLGVPDLPESEPVEMRQFVGCNRNIQVDSRHIDMADEFLAN 1564  
 4050 QFTGNCSEETACFPNRCRGGSCD---PIGNTFISCKAGL----- 4089  
 1565 NGTVPGCPAKKNVCDSNITCANGGTCVANOMDAFSCCEPLGFGKSCA-OEMANPOHFLGSS 1623  
 4090 --TGVTCEDDVDECEBCEBCEGSGSCVNLFGSPCNCTPGVYGCGLRPVVVPNIQAGHS 4147  
 1624 LVAMH---GLSLPISQPMVYLSIMFRTQADGVLLQALTR---GRSTITLQAREHVMLSV 1677  
 4148 YVGBELIGLINVLFVFTLVLPF-----VPRKKVPRKNYSNNNTLLVODPATALLH 4201  
 1678 EGTGLQASSLRLBERANDGDGMHQAOLALGASGPGHAILSFYQGOARABGNLGPRLHGL 1737  
 4202 KSNIGIPRSLA-----GGRNVYQEVGPQVPAVPAVTPPQFQDSRNLKGDAL 4254  
 1738 HLSNITVGGIIPGPGYARGRGLQGVRSVDPTEGVNSLDPSHGESINVEQ---CS-- 1792  
 4255 ---GG-----EPOELSTFPEBSPRLITLARGVVCSVA 4284  
 1793 ---LP---DPCDSNPCCANSYCSNDMSYS-----CSC----- 1819  
 4285 PNLRAVSPCRSD---CSIRKNGWDTGSEKKGADDTGSEVTCFANSNKGNSNEVSLANSFQ 4341  
 1820 ---DPGYGDNCTNVCDLNPCEHSGVCTRKS-----APH-GYTCBEPN-YLGPY 1865  
 4342 SDSGDNAHYMD---TSDMWPGARLSDIBEMFVYSQDGAVHQSGTRLESDYILGKY 4397  
 1866 ---CETRIDQPCPRGMWGHPTGCPNCVSKGFPDCKNTSGECCKE 1910

Db 4398 DISSETPPHEEFLSODQLPPPLP-----EDFPRQ 4428  
 Qy 1911 NMYRPSPTCLDCPTGSLSRVCDPBDGCK-----FGV--IGRQCDRC-D-NP 1960  
 Db 4429 YEALPSPQSPTSL-----TGTMSPDCKRRPRFRPSQVLPPLPGLSTDLGSPSSCDPST 4482  
 Qy 1961 FAEVTTNGCEV 1971  
 Db 4483 FAVSMNOGTEV 4493  
 RESULT 9  
 Q67F4 PRELIMINARY; PRT; 2920 AA.  
 AC Q67F4;  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE HMR-1B protein precursor (W02B9.1b protein).  
 GN W02B9.1 OR HMR-1 OR W02B9.1B.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N2 Bristol;  
 RA Broadhead I.D., Pettitt J.;  
 RT "A Caenorhabditis elegans neuronal classic cadherin, HMR-1B, is  
 required for axon fasciculation."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9069613; Pubmed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Leonard N.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY REPRESENTATIVELY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC EMBL; AJ307058; CAC38842.1; -;  
 DR EMBL; Z82064; CAD27611.1; -;  
 DR EMBL; Z82093; CAD27611.1; JOINED.  
 DR EMBL; AL032638; CAD27611.1; JOINED.  
 DR EMBL; AL032638; CAD27619.1; -;  
 DR EMBL; Z82093; CAD27619.1; JOINED.  
 DR EMBL; Z82064; CAD27619.1; JOINED.  
 DR EMBL; Z82093; CAD27620.1; -;  
 DR EMBL; AL032638; CAD27620.1; JOINED.  
 DR EMBL; Z82064; CAD27620.1; JOINED.  
 DR Wormpep; W02B9.1b; CE30357.  
 DR InterPro; IPR000152; Aex\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00028; cadherin\_13.  
 DR Pfam; PF01049; cadherin\_C\_term; 1.  
 DR Pfam; PF00054; laminin\_G\_1.

DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 13.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00282; LamG; 1.  
 DR PROSITE; PS00010; ASK\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 8.  
 DR PROSITE; PS00268; CADHERIN\_2; 13.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00025; Lam\_G\_DOMAIN; 1.  
 DR PROSITE; PS00025; Lam\_G\_DOMAIN; 1.  
 KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 POTENTIAL.  
 SQ SEQUENCE 2920 AA; 323910 MW; 5B0D4C8BD24694FB CRC64;

Query Match 7.5%; Score 1164.5; DB 5; Length 2920;  
 Best Local Similarity 24.3%; Pred. No. 7.3e-64;  
 Matches 470; Conservative 279; Mismatches 745; Indels 437; Gaps 74;

Qy 176 NTAPOQPPSYQATYVBNQPGTVPASLRATIDPDGBAGRLLEYTMDALFDSRNO----- 230  
 Db 974 DNSPQERPSYETSVSREAVGTSVITVPFADNDAAHAEITYSL--IDITAGEHQND 1031  
 Qy 231 --FFSL-DPVTVGAVTAAELDRETKSTHVFRTAODHGP--RRSALATLTIL----- 278  
 Db 1032 LDFEFLVNRSGEILILPIPKTKQ-PIFNVIADNDGIPALQSSAQVTLNVLDKQKA 1090  
 Qy 279 ----- 278  
 Db 1091 PKQTSPOCKPITVDENVELINKVILCRAVSSGDSRNSDVIYKLTAGSGPKNAESKPR 1150  
 Qy 279 -----VTDTDHPV 288  
 Db 1151 QNKKEKNGEWEVYIMGLDYEQVNTTLITATDMSRVASTTPVYEVADVNDVVPQ 1210  
 Qy 289 FEQDEYKSL-----RENTLEV--GYEVLTVRA--TDGDAPNNAILYRLLEGSGSPSE 338  
 Db 1211 FTVDLFTGIDEMTPNHELEKTKNGKPIYTVVAIDTDSGPN--EWHRIYVEANGAEETK 1269  
 Qy 339 VEIDPRSGVIRTRPVNREAVESQVTVZASDQGRD-----GRSTTAANFLSVEEDN 393  
 Db 1270 HFRIDELTGEIIFPNKEFPREKIDMYILTVESDRSVSLPGANGNKONVKQIYINDN 1329  
 Qy 394 DNAPOSEKRYVQVREBVTPGAPLYRTASDRDGSNAVVYSIMSGNARQGYLDQT 453  
 Db 1330 DNAPSEKRYVGRKESBGBHDVITIKAHLDHSLKRYTL-IGAGGRIPFGVRTDS 1388  
 Qy 454 GALDVSPLDVETTYEYTLRVRAODGRRPLSNVSGLVTVVOVLINDNAPFVSTPEQAT 513  
 Db 1389 GTIFVEPELPDFEASDQYHLVLASDGRNATN-----YIHIEDVNDNAPQEQOKYAT 1444  
 Qy 514 VL-ESVPLGLVLAHQALDADAGD-NARLEPRLAGVGDPTTINNGCWISVAELRE 571  
 Db 1445 VIEEDVDLPKLVFNHATDADDERKSRIVRLRQGDDEVRIKYSGETELVALDRD 1504  
 Qy 572 E--VDYFSGVREADHGTALTAASVSVTVLVDVNDNPTFTQ-----PEYTVRLNE 622  
 Db 1505 PRAGVPSNMFFVQALDDDNGNLGVADQVNVNRDINDNSPIPELPLFGIENRRPHSD 1564  
 Qy 623 AAVGTSVTVSAVDPAHSVITYTSGTNRN-----PSITSQGGGLVSLALPDY 675  
 Db 1565 ---GVFMDVQARDPDDPTTANANLEYGVNRKLINGESVFRIDNTGKIFAMRSLDEBI 1621  
 Qy 676 KLEROYVLAVTASD---GTRQDTAQIVNVNTDANTHRPVPQSHYTVANNEDRPAGTTV 732  
 Db 1622 SSERETILVRANDGVPSRSGFANVTIKVTDMMNANAPFEETIRYEGSVIERAPLGAAM 1681  
 Qy 733 LISATDEDTGENAR---ITYFMEESIPOFRIDAD-----TGAVTVQALDYEDQVSY 782  
 Db 1682 SFSAPFAD--EAKONVFTYQLESBSDYFYVTTDSDKQSVGVLRVQPLDYED----- 1734



```

OY 783 LATARDN---GIPKSDTTY-----LELIVDVNDNAPQPLBDSYQSGSVEDVPFTSV 834
DB 1735 --VTQBDGPHLGI-RVSDGRHDAEAHVVALVDRNHAPH-IHGATEHHRVEDVPFGTST 1790
OY 835 IQISATDRDSGLNGRVFTYFQGGDDGDFIVESTSGIVTRLRDLRENVAVYLAAYAV 894
DB 1791 GRYTATDRDAGDTAR--FRINQSDPKRQFTIDQ-DGLRVAVHTDREDILAYNYLILEAY 1847
OY 895 DKGMPPARTPMETVTVLVNDNPPVFEDDEFDV---FVEENSPIG-LAAVATVATDPD 949
DB 1848 DMSNNIGR--QWVAVYLQDVNDNGP---EPYVPRPCIFRRETPVNOQGTCEIRATDRD 1901
OY 950 EGTNAQIMVQIVEGNLPEVFOLEIF-----SGELTALVDLDYE--DREYVLL--V 995
DB 1902 TAEFPPPTMEVSPSKYSQYILAVIFNANGDNGSMITPLQDFRERAPVQKLELPL 1961
OY 996 IQATSNAPLVSATVHVRLDRNDNPEVLGNFELFNNTYTNSSSPFGAIGRV--PAH 1052
DB 1962 ILADRAGRNRBASVHYIIGDLNDNTWHDGRMTIHNVSYGLRKET---VIGRVYVDDAD 2017
OY 1053 DDDIDSLTYS-----FERGNEISLVL--NASTGELKLSPL--DNNRPLEAIVSVLV 1103
DB 2018 DMDLGDK--TFSWKDSRPFELSDKGSITWAGEMAAGTYMSANVHDNADDEDAVGVYTV- 2075
OY 1104 DGVHSVTAQCALRVITITDEMLTHSTITLEDMSPERFL-----SPILGLFIOVAAATL 1157
DB 2076 --IYNAVPOIAIDNGQSVOLLTAETPLQL---PDDFIRADNSQGLNDTTRKQEMTAVM 2129
OY 1158 ATPRHVVVFNVO-----RDTDABGHIILANLSVGPGRPGGPPFLPSBLOLRLYL 1211
DB 2130 GGDVAV--VDFSVQVQVGLATLQTRDVP--VLAVRFN-----ARSTYRDTMOQNGLIA 2178
OY 1212 NRSLLTISAQVLPFPDDNICLRPCENTMRCVSLRFPSSAPFLASSVLFRRPHVVG 1271
DB 2179 HRADLQKRLNVAIVGVIGIDMCKFTQCD--AGQTLMSADYDGIIVVSANSTVIVGNATSR 2236
OY 1272 LRCRPPGTPGYCETEVDLCYSRPGPHGRKSRSEGGVTCGRDGYGHEHVSABSRG 1331
DB 2237 DDCTCP-----VNRAPPA 2249
OY 1332 CTPGVCKNGGTCVNLVVGFKCDSPSGPEKPYCVVTRSPRAPHFTRGLRORFHTL 1391
DB 2250 CQHSICNDGVCHN--TNPGFCECNRNDGLKARCCQTTTRSFGNGRPAWYKMPACTSLNI 2308
OY 1392 ALSPATKRDGLLTVNGRF---NEKH---DPALEVYQEOVQULFSAAGSTTVVSPV 1443
DB 2309 SFSFMTTOSDALLFYNGPLLETLRNTHIEYSYIFLOLRGGRISLEVSNNQSGRSLEVA 2368
OY 1444 PGVSDGQWHTVQLKYNNKPLLGQGLPGSPSEQVAVVTVYDGCCTVVALRP--GSVLGNY 1502
DB 2369 STALNDGTWHDISVN-----OEGKRVELVNDNC-----RFLGAGADS 2406
OY 1503 SCQAQGTGSGSKSLDLTGPLLGVV---PDLSPESFVRMRQFVGCNRLQVDSRHID 1557
DB 2407 SCRRLVLPDDDERLNTVTPVQIGGLAPLSGGDYQTLF--RAGLNGCVRLNANVGDYD 2464
OY 1558 MADFLANNGTVCPCAKKNCVSNCTCHANGCTCVNOMDAFS-----CECPLEGCG 1606
DB 2465 LATPAFEONSEKGCYLMGATCDNSNVDSILNHCIH--GDCFADVQSGAMVAKVCVCPGNGG 2523
OY 1607 KSCAEMANPOHFLGSLVAMHGLSLPISQPYLSLMPFTRQADV---LLOAIRRGST 1663
DB 2524 ARCERRMEMIOPAQGAFLIYSRIAPF--EQVSDILFLISGVNAPAPLBSGTDSSQSY 2582
OY 1664 ITLQUREGHVMSLV---GTG-----LOASSILRERPGANDGWHHQAALGASGG 1711
DB 2583 VSTINLESQNGVTAAGKFDIGTGRRARQELVSEVLL--KENASYM-----LOFTRN 2633
OY 1712 BEHALTSPD---YQOQABENLGPRLHGLHSNTTVGIFPPAGVARGPGCL----- 1762
DB 2634 PFRASLSDINAVTAVTQIDKG---EPFSLQVNOITLG-----TOGQNKFGQSGCTGYRW 2684
OY 1763 -----QGVAVSDTPEGVNSLDPHSGHSINVEGCSLPDPCDNP---CRANGYCSND 1811

```

```

DB 2685 SKONLPLRKGAMDNESIVSISNMAG-----VQGGDLRTTCADLPAGYCGGSPVGVDF 2740
OY 1812 WDSYSCSDPG 1822
DB 2741 WKGPCTCNDG 2751

RESULT 10
O8MY75 PRELIMINARY, PRT, 1959 AA.
ID O8MY75;
AC O8MY75;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
GN pfl-cadherin.
OS PflCAD.
OC Pteryodera flava.
OC Eukaryota; Metazoa; Hemichordata; Enteropneusta; Pteryoderaidae;
OK NCBI_TaxID=63121;
RN (1)
RP SEQUENCE FROM N.A.
RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Sato N., Humphreys T.,
RT Zhang S., Tsukita S.,
RT "A novel amphioxus cadherin that localizes to epithelial adherens
RT junctions has an unusual domain organization with implications for
RT chordate phylogeny."
RT Evol. Dev. 0:0-0(2002).
CC - FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY REPRESENTATIVELY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AB075368; BAC06837.1; -.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C-term.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01043; cadherin_C-term; 1.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 8.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00282; Lame; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS50268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;
KW Glycoprotein.
SQ SEQUENCE 1959 AA; 214791 MW; F778D75002F084CF CRC64;

Query Match 7.1%; Score 1108.5; DB 5; Length 1959;
Best local similarity 24.0%; Pred. No. 1.3e-60;
Matches 445; Conservative 252; Mismatches 703; Indels 455; Gaps 72;

OY 176 NTAPQFPQPSYQATVPENOPAGTVPASIRAIIDPDGEGAGLEVTMDALFDSRSNQPSLD 235
DB 216 DNABEPFPECEVTVPEISPDALVAIVHAVDPDAGKNGKMYTDT-----NDMTSVH 270
OY 236 PVTGAVTAAEHLDEKTSVHFVTAODHGMPPRSALATLTIVTD----- 282
DB 271 PDTGVIITIGPLQAKRYTLTVF---ARDRGKPSLEASTPGRVITTVPVQSLTFSSILG 327
OY 283 -----NDHDPVPEQGGYKSLRENLEVGVEVLTVRATDDGAPPNAILVRLLEGSGGSP 336

```

```

Db      328 PGVILLDDDE--ELPEIKVSLMEDAPKGSVTTITKPSNDKSSQ-----FRPLANNMNP 381
Qy      337 SEVEEIDPRSGVIRTRGPVDRREVEESYQLTVEASDQGRD---GPRSTAAVELSYEDDN 393
Db      382 R--FHVQYTGDIYLRDLIDAHPTIEINVGITKDNDRPLRBPQKXIIIT-IPSET 438
Qy      394 D-----NAPOFSEKRYVVOREDVTPGAP---VLRVTASDDKGSNAV 433
Db      439 DKPAPNQVVKPRIRRAKSPSFDPSPYEGTVEEDVLASPVPAFVPRRILITVANGPSPN 488
Qy      434 VHSVIMSGNARGQFYLLAAGALDVSPLEVTETK-----EYTLRVRQDGRGRPLSVSG 489
Db      499 IORIIQAEFRREDPAIGLTSGLISRKHLREDNKGQVREFTVA-----SSAG 548
Qy      490 LVTVOVL-----DINDNAPITFVSTPFOATVLESVPLGLVLAHQALDAD---AGDNARLEY 542
Db      549 TTSVEVLKIKDVNDNAPQPPYPYIGSPENNGAPQVMTIQMEDLDDNEGMLVRY 608
Qy      543 RLA-----GVGHDPFTINNGTMTISV-AAELREEDFTSPCYEARDHGTPTLTASAV 596
Db      609 SILMAONSVGOAL-PAIDSESAVITTOVGNLDRERIPSYDIYKAVDGG--ANTGVTA 665
Qy      597 SVTVLDVNDNPPFTQPEYTRVRLNEDAAVGSVTVTSANDRDANSVITTOITSGENTRF 656
Db      666 TIYVSDENDNPPPEPOAKSFVEISENVPDSSVFKVLAIDRV-GINSYISISNGVEKF 724
Qy      657 SITS--OSGGGLSVLALPLDYKL-EROYVLAVTSADGTRODTAQIVNVVTDANTHREVFQ 713
Db      725 RIETDPDSGAIIVAKKIDPELINESKVLITBVDKNTATTELDLRVWDANDVPSFS 784
Qy      714 SSHYTVANVEDRPGCTTVLISANDEDTGEN-----ARITYMEDSIPQFRIDADTG 765
Db      785 QPRVLAHPEDPAKMTTIGVAAATDRKSEKSEIEFEVAVRPFPPSSFR--FVVGANG 842
Qy      766 AVTQAEID---YED---QVSYTLATARDN-----GIPOKSTTYLRLIYNDVNNAP 813
Db      843 SWAIKDLREPYEDDPQOIRENYITVYATIDNVHGGGERANTVTTFTVISITDNNAP 902
Qy      814 QPLADSQGSVYEDVPPFTSVLQISATDRSGLNGRVFTYFGGDDGDFIVESTSGIV 873
Db      903 EFL-PGYTLVPEMTPGAYEAGVEIIRDDDSNGPP----- 939
Qy      874 RILRLRENVAAQYVLAIVAVDKGMPAKTPEMTYTVLVNDNPPVPEODEPVPVEEN 933
Db      940 -THRVIDDGILSR-----FDV---S 956
Qy      934 SPIGLAVARVATADPDGCTNAQIMYQIYEGNIPEVFOIDIPSGELTALVDLYEDRDEYV 993
Db      957 SPIC-----VEDOL-RVSLLEEDREATPPY-----YI 983
Qy      994 LVIGATSAPLVSRATVHVRLLDRNDNPVLGNFELFNNTVYNNSSSPGGAIGRVPAHD 1053
Db      984 PVLVSDKHGVSATQTLVTIGDVNDNPPSPATKITLVYSY---KGOIPRAIGEVYAAD 1039
Qy      1054 PDI--SPSLTY-----SPKGNELSLVLAASGELKLSALDNNRLEALIMVLYV 1102
Db      1040 EDIVKGDISTYVTPPIPOFSDVDQTOGIIIEBTKAGTYDKVKEVDEKMPVICVIV 1099
Qy      1103 SDGVHSTVAQCALRTVITTDMLTHSTLRLEDSPERFL---SPILGLFIOAVAAATLAT 1159
Db      1100 -----EYKDIPEBAVSSSGSVRL-LSAEKFTIGTSMKDRFDLIDVLGA 1145
Qy      1160 PRDHVVENVQRTDAPGHIILNVSLSVGQPPG-----PGGPPFLSEDLQERLY 1210
Db      1146 KKENIDIPSV-----INV-----PGEENQDVMSAHGSPYVPEKMTQTL 1187
Qy      1211 LNRSLTLTISAQVLPF-----DNICLAREPCENYMACVSLARDSAPRTASSSV 1261
Db      1188 LAKDV---TESQLGVSFGKVPIDMCLDRGIC-ESSCTNYFNAI-----TDPITIIDGAD 1237
Qy      1262 LFRPHVVGGLRCRC-----PGFPGDYCETEVLDLYSRPGCPHGRCSRREGVYCLCRD 1316

```

```

Db      1238 TVSVTTIORACVCGAKIAPRG-----PCASIPCLANGQCIDTBSGYCCQDE 1286
Qy      1317 GYTGHECEVSABSGCTPVCNKGCTCVNLVGFRCDCPSGDFFKPYCQVTRSEPAHS 1376
Db      1287 KXGG-----PNCEDISRGQNS 1304
Qy      1377 FTTFRGLRQRFHTLALSPATKERDGLLYNG-----RNEKHDVPALEVIOEQVLTSS 1431
Db      1305 FAMYPELSCBETKXISIEFLTYSRDGIILYNPIVPIIDLGQDFWALIELVGKPELTMD 1364
Qy      1432 AGESTTVPFPVGGVSDGQMTVOLKYNNKFLGQTLPGQPSRQKLVVVVDDGDTGV 1491
Db      1365 LSGMLPLTIDSP-PLNDEKMTLET-----IRSKHVEFLDHCENAV 1407
Qy      1492 ALRF--GSVLGNVSCAAGTQGSKKSLDTLTPLLLGVP-DLPSPFVPRM-----QFV 1543
Db      1408 IYETERSTVSDSNCNCAATDPTTGTRFLVNVTFPLQGGIDMDSTYTPYTFNPARQGDV 1467
Qy      1544 GCMRNLOVDSKRIIDMADFIANNGTVPGCCAKNVCDSN---TCHNGCTCVNQMDAPSCPC 1600
Db      1468 GCIRNYIDGKYDLDGTPGREHNSQGCPTDILCYDGVVYCKNGLCVTESASVYTCV 1527
Qy      1601 PLGFGKSCAOMANPOHFLGSSLVAMHGLSLPISQPM---YLSIMFTRQADGVL--LQ 1655
Db      1528 DEGWGEACDEBLPS-YDPLSDSYVYIDLABSLVDKTVSDYQIMVGNTOBEDALLMHQ 1586
Qy      1656 AITRGSTITTLQLRBGHVLAVE-GTGLQASSLRLEPGANDGDWHHAQALGASGPGH 1714
Db      1587 NVD-GEYITTMKKEGNAPFIBYNLGDPKMYSLRVANVTYINDGNVHTVL----- 1635
Qy      1715 AILSPYGGQRAEGNLGPRHLGHLNITVGGIIPGAGVAVAGFRGCLQGVRSVDPREG 1774
Db      1636 -----LRYGNIFRFVDG--GG-----GVTVKAAQGI 1661
Qy      1775 NS---LDPGH---GESINVEQCSLPDPCSNPCPANSYCSNDMDYSQSCC-----PGY 1823
Db      1662 FKRPIDPSSVYGVANVSLSEKRVLGD--DYGCAKADALISYSYIGFDEIDGTLAVPR 1719
Qy      1824 YGDNCTNVCDLNPCEHGSVCTRKPSAPHGYTCBPEN-YLGPYCTRIDQPCR 1876
Db      1720 VERDC-NICLSFICIAPLVCVPTPSF---YCSCPEDLVFRGAC-VKSGGPAQ 1769

RESULT 11
ID Q19319 PRELIMINARY; PRT; 4307 AA.
AC Q19319; Q19785; Q21606;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical 479.0 kDa protein F25F2.2 in chromosome III.
GN F25F2.2.
OS Caenorhabditis elegans.
OC Rhabditidae; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Ainscough R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Z35662; CAAB4721.1; -.
DR EMBL; Z35599; CAAB4721.1; JOINED.
DR EMBL; Z34802; CAAB4721.1; -.
DR EMBL; Z35599; CAAB4661.1; -.
DR EMBL; Z35662; CAAB4661.1; JOINED.
DR EMBL; Z34802; CAAB4661.1; JOINED.
DR EMBL; Z34802; CAAB4338.1; -.
DR EMBL; Z35599; CAAB4339.1; JOINED.
DR EMBL; Z35662; CAAB4339.1; JOINED.
DR HSP; P00740; 1EDM.
DR WormPep; F25F2.2; CE00961.

```

DR InterPro: IPR002126; Cadherin.  
DR InterPro: IPR006209; EGF like.  
DR InterPro: IPR006210; EGF.  
DR InterPro: IPR001791; Laminin-G.  
DR Pfam: PF00028; cadherin; 23.  
DR Pfam: PF00008; EGF; 3.  
DR Pfam: PF00054; laminin G; 1.  
DR PRINTS: PR00205; CADHERIN.  
DR SMART: SM00112; CA; 28.  
DR SMART: SM00181; EGF; 4.  
DR SMART: SM00282; LamG; 1.  
DR PROSITE: PS00232; CADHERIN\_1; 9.  
DR PROSITE: PS00268; CADHERIN\_2; 23.  
DR PROSITE: PS00022; EGF\_1; 4.  
DR PROSITE: PS0186; EGF\_2; 2.  
DR PROSITE: PS00025; LAM\_G\_DOMAIN; 1.  
DR Hypothetical protein; Calcium-binding; Cell adhesion;  
KW EGF-like domain; Glycoprotein; Repeat; Transmembrane.  
SQ SEQUENCE 4307 AA; 478963 MW; 28536C08264E151 CRC64;

Query Match 7.0%; Score 1092; DB 5; Length 4307;  
Best Local Similarity 21.6%; Pred. No. 5.8e-59;  
Matches 417; Conservative 248; Mismatches 641; Indels 628; Gaps 58;

QY 232 FSLDPTVGA-VTTAEE-----LDREKSTHVRV-----TQDHGMRRAAL 272  
DB 2410 FSLAFTSSNSTSKAMPVAIDRKTGRHVSRLVNYHRDKRYQLPVED--ATGRRAF 2467  
QY 273 ATLTLVDTNDHPVPEOQYKESLRNLEVGVEYLTVRATGDGAPNANILYRLLEG 332  
DB 2468 STLTSLVINDKPPFVLPFYSTSLSEAKGSDTYMWSAIDDD--ENDTLEYSLDD-- 2523  
QY 333 GSPSEVFEDIDPSGVIRTRGPVDRREVSQYLVASDOGRDGGRRSTTAAVFLSVED 392  
DB 2524 -GSESGPFVHPRQGITVAKKLEHAGVTLSTIKATDSANPFAHATTVEV--NIASE 2580  
QY 393 NNNAPESEKRYVQVREDVTPGAPVLTARTASDRDGSNAVHYGISMGNAQO---FY 448  
DB 2581 SVKVPPESSHVLPFVMEADVGNVIGVQOMETIDE--IRFTVPSBSBSDFLPS 2637  
QY 449 LDAQGALDNVSPLDVETTKETYLTVRAQDGRPLSNYSGL--VTVQVLINDNAPIV 506  
DB 2638 VRSSTCKIVKSSLDERRKQKMTIRADAG-----GHATTYVVDIGVNDNAPAH 2692  
QY 507 STPPQATVLESVPLGLVLAHQALDADAGDAALELYRLAGVGDPEPTINNGTWISVA 566  
DB 2693 GYERPTISEDAAVGTSTVTFSMDRDPSGRIRPSL--VEENPYFDLNNMSGMLTVAS 2750  
QY 567 ELDRBEVDYSPGVEARDHGTPLTASASVSATVLDVNDNNPTFTQPER----- 615  
DB 2751 QLDREKIDHYKLIARATDEG--GFNTDLPFTIVSDVDSPPQFEKEEFNIDLHPSTSP 2808  
QY 616 ----- 615  
DB 2809 ILHFSIKDDDLSPNNVSQPIPKGNEBGVFWIDSNNDLKKRPEIYENKQYQLKTAFD 2868  
QY 616 -----TVRLNEDAAVGTSV-----VTVSAVDR 637  
DB 2869 GVFEETSTVKYNLKSSKSDDIRCPBANKTVILAENSKGTVVLGESSLLGPVNTFLSDN 2928  
QY 638 DAH-----SVITYQI----- 647  
DB 2929 DSNVPPVNFNGIVKVESDQIDYEKNQOLEPFRLLTIQNSEVCKELLTVIENLINDNP 2988  
QY 648 -----TSGNTN-----RFSITSGSG-----GLVS 668  
DB 2989 KIIKELKVSIDENLPTSDARQYITRIYAEDADPEIKFRWDDCGELFOJDDINGVTT 3048  
QY 669 LALPLDYKLERQYVLAVTASDGTRODTAQIVVAVTDANTHERPVQSGSHYTVANVEDRPA 728  
DB 3049 VVPEFLDESEILGFNLTVVASDGEFEDKATILVLIQNDNAPLFEKSTVSMKMESESIG 3108  
QY 729 TTVVLISATDEDTGEMARITTFWEDS-IQGF-RIDADTGAVTTOALDEDQVSTLAI 786

DB 3109 YELAHFRASGSD--QNETLEYLTKPSDVTSPFNINAQGTGILTLAFLPETLSALKLTV 3166  
QY 787 ARDNGIPQSDPTTYEILVNDVNDNAPOFLRNSYQGSYVEDPPTFSYQISATRDGSL 846  
DB 3167 AKDSVPLEETRAQVEISVMDENDNAPFEKRYKQYKYNKSKSEKYLTVALLDVDSRH 3226  
QY 847 NGRVFTYFGGDDGDD-----FVSESTSGVETRLRLDRNVAQYVLRVAVNDKMPBA 901  
DB 3227 FCAVSEYEIYSETTDPVLPFAINS--NGDVLISQSIDYKIKKYNLKVIAKDGGRPL 3285  
QY 902 RTPMEVTVTVLDVNDNPPVPEQDEPDVVEENSPGLAVARTALDPPBGTAAQIMYQIV 961  
DB 3286 RSEALLEIHEDENDHAPTFDCCNMTALVQSEBALGHRLLKRSVDLDGPKKGAFTVEI 3345  
QY 962 EGNIPFOLDIFSGLTALVDLDYEDRPEYLV-----QATSAPVSRATVSRLL 1014  
DB 3346 QGDGAKSKYN-EKLELTAKLETRKDKTLLTYIADQYKTTDCEP-----TIFIKOT 3400  
QY 1015 DRNDNPVLGNFELLFNNVYTNRSSPPGALIGRVPAHDPIIDS--LTSFERGENELS 1071  
DB 3401 SR--HAPTMKPKIQINTL-----QNELPEGITGRKASDEDEBDQNGLLRFLVBSIOS 3454  
QY 1072 -----LVILNASTGEL-----KLSRALDNNRLEALMSVLYSDGVHSTVAQCAL 1115  
DB 3455 PRAQVESRSTHLPFVDPDNTGDIWSDHSITQGLHT-----FNVTVDSKENTSYVEV 3507  
QY 1116 RVTITDEMLNHSITLRLDEDSPERFLPPLGLFIOVAATLATPPDHV-----VFN 1168  
DB 3508 HTTSLDNDVIDHAYSIRIRMSVDPEMKRYAFERRIISHHLNLDSSIQILISQVAVPS 3567  
QY 1169 VQRTDABGHIILNVSLSVGQPPGPGGPPFLPSEDLQRL-----YLNRSL 1216  
DB 3568 TESERRSRNSMEDVEILMTAQRLGRLG--YIKPDHYSIRLKNDQNNNDQSQRMYQI 1269  
QY 1217 TAISQRLVPPDNICLAEP-----ENYRCQSVLRPSSAPFIASSSVLFPPIHV 1269  
DB 3626 TEMCT-----TGCLNGCREVIELIEDSWTKYSTDFVFSFHSRSA----- 3669  
QY 1270 GGLRCRCPGFTGVDCETREVDLCYSRPGPHGRG--RSREGYVCLCDGYTGEHCEVSA 1327  
DB 3670 ---QCLCDGGRGRCRVEFNQCAKSPCEQMLCIPSHNSTYBEVCLQMEGDCSPS 3726  
QY 1328 RSGRCTPVCNKGTCV---NLVGGFKDCPSGDPEKPYC-QVTRSFPASHFTTFRGL 1383  
DB 3727 ---CONDGCLAEALSVGG-----DGFEISLSNEIETR----- 3758  
QY 1384 RQRFHTLALPATTERRGILLNCRFRBKNDPVALAVIQOVQLTFSAGSITTVSPFV 1443  
DB 3759 ---MELIEBLKTTTHNGIIMWS---RGKDPFHLRLVNGTPEYHMAAGTGTIVTS-- 3808  
QY 1444 PGVSDGQMTVQLKYYNKPPLLGQGLPQGSBQKVAVAVTVGCDYVALR-FGSVLGNV 1502  
DB 3809 KTSVVDGQMRIALS-----RRQRRTMTYDDELOARSPISGTVLNL 3852  
QY 1503 SCAAGTGGSSKKSIDLTPLLLGVPDLPSFVNRMQVYCKRNLOVDSRH-----D 1557  
DB 3853 HRYSGKVLGAK---VDDGELTDGVS-----ACFEITISVDGMKVLKTRQG 3894  
QY 1558 MADFIANNQVPGCPA-KQNTCDNSTGNGGTQVNMOMA-FSCBPLFGGKSCAQEBAN 1615  
DB 3895 MGLF---GAQPGSALTSPPCNDLPQHAQGTCSQGSNHRKCEPSSRYSGNVEID-- 3948  
QY 1616 POHFLGSSLVAMHGLSLPISQPMWLSLFRTRQADGVLLQAITRGRSTITLLREGHVNL 1675  
DB 3949 ----- 3948  
QY 1676 SVBGTGLASSLRLEPRANDGMHQAOLAGSAGGHAILLSFDYGOQARABNIGPLH 1735  
DB 3949 ----- 3948  
QY 1736 GLHLSNITVGGIPRAGVGARPGCLQGVAVSDTPBGSVNLDP9HGSIINVQCCSLPD 1795



DR PROSITE; PS50268; CADHERIN 2; 13.  
 KW Calcium-binding; Cell adhesion; Glycoprotein.  
 SQ SEQUENCE 1881 AA; 209226 MW; 109F3F517BAF214 CRC64;

Query Match  
 Best Local Similarity 5.8%; Score 906; DB 5; Length 1881;  
 Matches 315; Conservative 176; Mismatches 413; Indels 308; Gaps 42;

QY 182 QPESYQATVP-----ENOPACTPV-----ASLRALDPDEGEAGRLTYTMDALFDSRS 228  
 DB 275 QPPEFLYQPVVRISEDAPICTEVLARMVSLSTKAVDGRGINNRIYGI-----SNGG 330  
 QY 229 NQPSLDPVGAVTAAELDRETKS-----THVRVTAQDHG--MPRSALATITLIVTD 281  
 DB 331 SELEIDRLKGLSTKQKLDREDSNTPINGAFILAEVIAESKIQAPASSIMETLIVTD 390  
 QY 282 TNDHDPVEEQEYKESLENTLEVGVLTVAATDGDAPPANILYRLLEGSGS----- 335  
 DB 391 VNDSEIPRRSDGVCGEIGENAG--ENTLAFIDG-----SINEVPDIDQKNGTFRSLH 443  
 QY 336 -PSEVFEIDPRSGV-----IRTRGP--VDREEVSYQLTVEASDQGRDPGPRSTTAVF 386  
 DB 444 PPSDIFFEVTPKRALNEATFGLRVKDPMSLDYERVELSTLVASEV--ESAGRTSTQIR 501  
 QY 387 LSVEDDNDMAPOSEKRYVQVREDYTPGAVLAVTASDRKCG--SNAVHTSIMGAR 444  
 DB 502 VVAVLDQNDNPFEPQPYVDIDVPRNVLAGVLLQIQATDSGSLGTEGVRYAMTGSIS 561  
 QY 445 GQPYLDAQTCALDVV---SPL-DYETTKETYLTVRAODGGRPLSNVSGLVTVQVLDND 500  
 DB 562 SFPLHDHAGVTLMASBSFVPRDEITIQKHTLSLEABDNGRGKRVPLI-LAVLDVND 620  
 QY 501 NAPIFVSTPPQATVLESV---PLGYLVTHVQAIDADGNALLEYRLAGVGDHPPTI 555  
 DB 621 NAPIFVDRKREYERLKENAFEFESP--IVVEARDSDLEGSPNSAVERRLGAHSDYFHV 677  
 QY 556 NNGGWSVAAELDREEDF-----YSGVETARDGTALTRASVSATVL 601  
 DB 678 DRRIGRLSV---REANDFERLESSGGSGDTRTIALITLAEADGSEPLTAQVEVTVYQ 732  
 QY 602 DVANDNPTFTQPEYVTLNEDAAVGTSVVTVSAVDRDA---HSVITYQITSGNTRNRSI 658  
 DB 733 DVANDYAVFLBSQYAIYIPEDTPSGLVLRVLTAMDGSGSPNNHVTTRIQGGD-GAFVI 791  
 QY 659 TSGSGGLVSLALPLDYKL-----EROYVLAVTASDGTROD--TAQIVNVTDANT 707  
 DB 792 GASTGEISITHGASILDENLLAPDALGSGSFYLEVPAFDGNGDQQLQSGCLVNT----- 847  
 QY 708 HRPVFGSHYTVVNNBRPAGTTVVLSATDEDTGENARITTFMEDSIPQ----- 757  
 DB 848 -----PVGETEYRLMATDPD--EGAMLRYYIDRSLSGKTEBEGALVK 887  
 QY 758 -----FRIDAGVTTQAELEDYEDQVSYTLATARDNGIPQSD--TLYLEILV 805  
 DB 888 LDDYDFAAALINBTNLKIALKLDREKTAETLACVDEVA-AEGDDQMANITFLKTV 946  
 QY 806 NDVNDNAPQPLRDSYQSGSYVEDVPPFTSVLQISATRDSGLNGRVFTYFQSG----- 857  
 DB 947 LDEDDNPKPKRPKYKHSIAENSQYGAVCTVVAEDADQ--NKTVKXSLGEGKGVLELH 1004  
 QY 858 -DDGDDGFIYESISGIVTRLRLDRENVAOVLAAYAVDKGMPARIPMEVTVTVLDND 916  
 DB 1005 VDDTGTIVRN-----RIDHEEYSMLNPSVRADPTGTPRPSFVEVFPQVQVDEND 1055  
 QY 917 NPPEFQDEDFDVEENSPIGLAVARTATDPDGTNAQIMYQIVGENIEVEFOLDIFSG 976  
 DB 1056 NNPFYVDSVNDYVYSENASGARIAIILADLDOSGDGRITTYILDRSSSEKXSIDBERG 1115  
 QY 977 ELRALVDLYEDREPYLVIOA-----TSAPLVSRAVTVHLLDRNDNPPVL-- 1023  
 DB 1116 ILRVAGALDREETAELVMAVEAWNDYGYLNGESRRAFKHILIHV--LDDNDNVPIQK 1173  
 QY 1024 -----GNFEILFN----- 1031

DB 1174 PSCGSMTEYHNINDPIVKLATDADPTNGNGQLGFDIYDPGSGIFYIQVSAQYAEIYS 1233  
 QY 1032 -----NY-----VNRSSSP----- 1042  
 DB 1234 RGPLKHLNGYTYLLEIVSDIGVPTNARBSVDICVDENHAFVFPVPSGNTVYKVEPT 1293  
 QY 1043 --GGAIGRPVANDPDIDSILTVSFE-----RGNELSLVILNASTGBIKLSRALDNNR--- 1092  
 DB 1294 TLGKPFQVAVYDEVDGENNAIVRYRLKMDTMGN-PRKFSLDKETGELSLAPIDREQQM 1352  
 QY 1093 -----PLEALMSVLY-SDGHSVTAQCALRVTI--TDEMILTHSITLLED 1135  
 DB 1353 YDLRIEAYDQGIPTPLSTVDLIYVADVVDNLPQFLIKESISINTEHMTPTGTERIRLPD 1412  
 QY 1136 MSPERFLPLLG 1147  
 DB 1413 TVDDYLDPLDG 1424

RESULT 14  
 096UL3  
 ID 096UL3 PRELIMINARY; PRT; 803 AA.  
 AC 096UL3; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein KIAA1812 (Fragment).  
 GN KIAA1812  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=21245130; PubMed=11347906;  
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 8:85-95(2001).  
 DR EMBL; AB058715; BAB47441.1; -  
 DR InterPro; IPR002126; Cadherin.  
 DR Pfam; PF00028; cadherin; 6.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 6  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS50268; CADHERIN\_2; 5.  
 KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;  
 KW Glycoprotein.  
 FT NON TER  
 SQ SEQUENCE 803 AA; 88205 MW; 15F2EDP00EFPAE9 CRC64;

Query Match  
 Best Local Similarity 5.8%; Score 905.5; DB 4; Length 803;  
 Matches 254; Conservative 104; Mismatches 326; Indels 41; Gaps 15;

QY 241 VTAAELDETSSTVFRVTAQDHGMPRSALATITLIVTDNHDHPVEEQEYKESLAE 300  
 DB 4 IRVALPLDEYTDVDRDFDLFANE-SVPDHFVYAKKVTILINNDKRPISQPLYNISLYE 62  
 QY 301 NLEVGEVTVVATDGDAPPANILYRLLEGSGSPSEVFEIDPRSGVTRTGPVDRBEV 360  
 DB 63 NTVGVISVLTVALTNDNAGTPGEVYFF-----SDDPDRFSLDKOTGLIMLARLDYELI 117  
 QY 361 ESYQLTVEASDQGRDPGPRSTTAAVFLSVEDNDNAPOSEKRYVQVREDYTPGAVPLR 420  
 DB 118 QRFYTLITIA---RDGGGEITIGRVIRINVLVDNDVPTFOKQAYGALRENEPSVTVLVR 173  
 QY 421 VVASPDKDSNNAVHVSIMSGNARGQFY---LDAQTCALDVVSPLDYETKE--VTLRYR 475  
 DB 174 LRAITDEDSPPNNQITYSIVSASAFSGSYFDISLIEGYVLSVRPLDYEOISGLIYLTVM 233

QY 476 AODGRRPLSNVSGLVTVQVLIDINDNAPFVSTPQATVLESVPLGLVLAHQALDADG 535  
 234 AMDAGNPL-NSITPVITIEFENDENDNPTFSKAPFVSVENIMAGTIVFLNATDLDRS 292  
 QY 536 DNARLERLAGVGHDPFTTINNGTWSVAELDREVDVPSGVEARDHGT--PALITAS 593  
 293 REYGOESIIYLSGASTQFRINASGEITTSLLDREKSEYIILVRAVDGVEGHNOKTGI 352  
 QY 594 ASVSVTVLVNDNNPFTQPEYTRNLNEDANAGSVTVASVARD--AHSVITYQTSGN 651  
 353 ATNITITLIDNDNHPFKAPYITINLVEMTPPSDVTIVAAVDPDGEENGLVYSIOPEN 412  
 QY 652 TRNRFSITSGGGGLVSLAL-----PLDYKLERQVYLAVATSDGTR-----QDTAQIV 699  
 413 --KFYSINSTTKRIKRTTHAMLDRENPDPEAEMLRKLVSVT--DCSRPLKATTSATVP 468  
 QY 700 VNTDANTHRPVQSSHYTVANNEDRAGTGVVLISATDDEDTGEMNATITFMEDSLPQ- 757  
 469 VNLIDLDNDPFTQNLPEFAVEVEGLPAGVSIQVVAIDDBEGINGVSRMPVGMFRMD 528  
 QY 758 FRIDADTGAVTQALDYEDPOVSYTALATARDNGIPKSDPTYLEILVNDVNDNAPQFLR 817  
 529 FLINSSGGVVTTEHLDREKIAEYQLAVASDAGTFRKSTSTLTIHVLDVNDETPTFP 588  
 QY 818 DSYQSVYEDVPPTSVLQISATDRDGLNGRVFYTFQGGDDGDPDIYESTSGIVRTLR 877  
 589 AVTVSVSEVDVPEFRFVVMINTGNDVGNALSYFTIGG--NVDGKFSVGRDAVVRVTV 647  
 QY 878 RLURENVAQVTLAAVNDKMPKAPRTMET--VTULDVNDNPPVQDEPDVPEVBNRP 935  
 648 GLDRETTAAVMLILEALDNGPVGRKRTGTATVPTVLDVNDNRPITFQSSYEASVPEIDP 707  
 QY 936 IGLAV 940  
 Db 708 EGHSI 712  
 RESULT 15  
 Q9VVG0 PRELIMINARY; PRT; 1820 AA.  
 AC Q9VVG0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative cadherin precursor (CG6445 protein).  
 GN CG6445.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=2019606; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Aroll J.F., Agbayani A., An H.-U., Andrews-Plannoch C., Baldwin D.,  
 RA Bailew R.M., Baau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Bouch M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Paolo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Leavitt P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier R., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Stylianakis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weasman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL ADHESION.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 DR EMBL: AE003524; AAF49351.1; -.  
 DR HSSP: P15116; INCU.  
 DR FlyBase; FBgn0036715; Cad74A.  
 DR InterPro; IPR002126; Cadherin.  
 DR Pfam; PF00028; cadherin.12.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA, 13.  
 DR PROSITE; PS00232; CADHERIN 1; 4.  
 DR PROSITE; PS00268; CADHERIN 2; 13.  
 KM Hypothetical protein; Cell adhesion; Glycoprotein; Repeat; Signal;  
 KM Calcium-binding.  
 FT SIGNAL. 1 14  
 FT CHAIN 15 1820  
 FT DOMAIN 36 128  
 FT DOMAIN 137 245  
 FT DOMAIN 254 365  
 FT DOMAIN 374 483  
 FT DOMAIN 492 585  
 FT DOMAIN 590 696  
 FT DOMAIN 718 815  
 FT DOMAIN 824 939  
 FT DOMAIN 955 1053  
 FT DOMAIN 1062 1170  
 FT DOMAIN 1170 1284  
 FT DOMAIN 1295 1394  
 FT DOMAIN 1407 1517  
 FT CARBOHYD 191 191  
 FT CARBOHYD 371 371  
 FT CARBOHYD 435 435  
 FT CARBOHYD 678 678  
 FT CARBOHYD 727 727  
 FT CARBOHYD 868 868  
 FT CARBOHYD 890 890  
 FT CARBOHYD 1040 1040  
 FT CARBOHYD 1111 1111  
 FT CARBOHYD 1112 1112  
 FT CARBOHYD 1233 1233  
 FT CARBOHYD 1294 1294  
 FT CARBOHYD 1419 1419  
 FT CARBOHYD 1452 1452  
 FT CARBOHYD 1557 1557  
 FT CARBOHYD 1584 1584  
 FT CARBOHYD 1610 1610  
 SO SEQUENCE 1820 AA; 200096 MM; 8D61B907FA15C155 CRC64;  
 Query Match 5.8%; Score 905; DB 5; Length 1820;  
 Best Local Similarity 25.4%; Pred. No. 9,3e-48;  
 Matches 350; Conservative 214; Mismatches 468; Indels 346; Gaps 55;  
 QY 180 Q9QPP-----SVQATVPBNQPGTGVASIALRDLDPBGEAGRLVETMDLAFDSRSNQFSL 234



```

Db      | 239 QDDPRTNAPYKATVNTENTPAGVSIITVYKAVDGDVCIIPREIPLSLE---DEPPGHELV 295
Qy      | 235 ---DPTGAV---TTAEILDRETK-----STHVRVTAQ---DHGMPRSALATITLV 279
Db      | 296 PFGPGRGTAVLQTTSEPLDRENAEILONGVYFYSIRATELIDGALPAHSLTRVITV 355
Qy      | 280 TDTNDHPVEQOEYKESLRENLVG----- 305
Db      | 356 TDDVDHQPFTSGPHFNVSITENLANGMPLPGLSI FVDDRDGMGNSRYEISLRDVFNARV 415
Qy      | 306 -----YELT----- 310
Db      | 416 FEVSPTESQGRTPVVVYVNLASRLDYVDPDLKPEFDLVASVKYERAKTVEIHL 475
Qy      | 311 -----VRATDGAPEPNANILLYRLLEGSGSPSEVFE 341
Db      | 476 ANDNAPVFDQGYRFTAAENI.PVDALIGHYKATDLSGEPGHVRY-VLKQFG---ADNFY 531
Qy      | 342 IDPRSGVIRTRGVDEBEVESYQLTVEASDQGRDPPGRSTTAAVPLSVEDNDNAPQFSE 401
Db      | 532 VNPETGAVYILKPLDYEKSSYSLTVVALDQO---REANANILFVGVTVDVNDHNPFS 587
Qy      | 402 KRYVQVREDVTPGAPVLRTVTAASDRDKS--NAVHYSIMSGN--ARGOPYLDAQTALD 457
Db      | 588 KEYSRTIREGALPEQOPVRAHADQPSQGNKRVKSIYSENSIAGNVFRIEPTGEIV 647
Qy      | 458 V---VSPLDYETTKETTLVRAQDGGAPPLSNVSGLVTVQVDIN-----DNA 502
Db      | 648 IQKASMDTE-RGEYELVVSATDFGIPLSNTTRVLVRVGISGNORPIFRGHQNMENL 706
Qy      | 503 PIFVSTPFPQATVLESVPLGVLVHQAIDADADNARLERLAGVHDPPFTTNGTMI 562
Db      | 707 PIIIGPSPYRASIPEKNAAGSNVTSVAHDPD-GDLSLRRIYGANDNEIDEISGLITV 765
Qy      | 563 SVAAELDRE-EVDFYSGVEARHDGTP-ALTASASVTVTLVDVNDNPPFTOPEYVRLN 620
Db      | 766 SPQARIDRDSNMMSFELIYNAVDSCTPIPETATTVYVNVKDINDRPPKEQNSYATYVS 825
Qy      | 621 EDAVAGTSVYVTSVAVDRDAHSVITTOITSGTNRNRSITSQSGGLVSLALPLDYKLERO 680
Db      | 826 ERTAVGESVLRVYAIKDLMSKLEYSMV-----GPVAATTKAG--VSJANRSNYRLOEA 877
Qy      | 681 YVLAVTASD---GT-RQDTAQV---VNTDANTH-----RP 710
Db      | 878 FRVDSQSGELFVNGTLRHDAIILITVGRDINAARVDEQVDTTEVTYVQSFQDTNP 937
Qy      | 711 VFQSSHYT-----VNVNEDRPAGTTVVLISATDEDTGENARITYF---MEDSIPQFRI 760
Db      | 938 VFRNTGWTSSRPVIDYKIKEMPIDALPILQADPVTRO--PITSFELIEPKQVDYFOY 995
Qy      | 761 DADTGAVTTOAELDYE---DOVSYTLAITARDNGIPQKSDTTYLEILVNDVNDNAPQL 816
Db      | 996 AERTGEVILKRLDYBALGDTGPEFELQYRA-NSADRQSTSVSRVNIIVENVNDNSRFE 1054
Qy      | 817 RDSYQGSVVEDVPEFTSVLQISATDRDGLNGR-----VYTTQGGDDGDDPFIYES 868
Db      | 1055 RNSYQATIIENRPHRVRRLDKDAVLANADERLGYHKIITSLOG--EHAMLFDINN 1112
Qy      | 869 TSG--IVRTLRLDRENVAYVLRAYAVDK-GMP--PARTMEVTVTLVDVNDNPPVFRQ 923
Db      | 1113 TTGEIIVASGQITDRERTRIQIQIKAESPRGFTPAKOSVVELQIEVDVNDNABEFTQ 1172
Qy      | 924 DEFDFVEENSPIGLAVARVATDPDEGTNAQIMQIV-EGNIPVFOLDIFSGELTALV 982
Db      | 1173 KKSSTVIPENAOIDSFVLQLEAVDADEGLGGEVRELVNVEGEANGLFKIDTKSGLISTR 1232
Qy      | 983 DLDYEDRPE-YVTVIQA-----TSAPVSRAT---VHVRLLDRNDNPPV-----LGNF 1026
Db      | 1233 NLTKGRAPHPYVILVRAQDNGNMPKOPTLSTDTDVIRIYIGVDSANDGVYFLSPRVQM 1292
Qy      | 1027 EILFNMYVTNRSSSPFGAIGRVAHPDP-----ISDSLTVSFERGN-ELSLVILNASTGB 1081

```

```

Db      | 1293 ANTENAVT-----GAPVFOVIASDPDDDESTPSCGTTTTRILIPDTFPDAEAFIDAHSGI 1345
Qy      | 1082 LKLSRALDNNRPLEAIMSVLSDGVHSAVTAQCALRVITITDEM.LTHSITLRLDMSR--- 1138
Db      | 1346 ITRQSMO--RFTKDMYRILILEVSDNGPKQKQSATRILOIA-----ILDVDHPRFA 1395
Qy      | 1139 -ERFSLPL-LGFLQAVATLA-----TPPHVVV-----FNQR----- 1171
Db      | 1396 READAGPLMSVAREEPACTIVGNPSALDEDDIGENAAIDYVITIDGNNBQLFTIERNNESI 1455
Qy      | 1172 -----DTAPGCHIINVS-LSVQPPGPGGPPFLPSEDLQERLYINRSLTAISA 1221
Db      | 1456 AILTKKPEIDRQVBSFTLTIKCLKLGBGYTFIDPT-----DRQDLSHRLNV----- 1505
Qy      | 1222 QRVLPEDN-----ICLRPCENYMRCVSVLRFDSAPPIAS--SVLFRP 1265
Db      | 1506 -RVLDDIDNLPRFQOPDPTVGIRINVPIDTVVTTLKASDADAAPVGLSIENTVFPV 1562

```

Search completed: February 11, 2004, 15:57:08  
Job time : 96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 19:27:01 ; Search time 11477 Seconds  
(without alignments)  
6189.944 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 15545  
Sequence: 1 MSPATGVLPTPPPLLL.....AGTDESSGSSEFLPFNFILH 2923

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Pgapop 6.0 , Pgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=csn2.1/USPTO.spool/US09916849/runat.11022004.143816.20951/adb.query.fasta.1.3079  
-DB=EST -QFMT=fastap -SUPPLY=rest -MIMATCH=0.1 -LOOPT=0 -LOOPTXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=PCO -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849 @CSN 1.1 9354 @runat.11022004.143816.20951 -ICPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST: \*  
1: em\_escba: \*  
2: em\_escbm: \*  
3: em\_escfn: \*  
4: em\_escmu: \*  
5: em\_escov: \*  
6: em\_escpl: \*  
7: em\_escro: \*  
8: em\_esc: \*  
9: gb\_esc1: \*  
10: gb\_esc2: \*  
11: gb\_esc3: \*  
12: gb\_esc4: \*  
13: gb\_esc5: \*  
14: em\_escfn: \*  
15: em\_escfn: \*  
16: em\_escfn: \*  
17: em\_escfn: \*  
18: em\_escfn: \*  
19: em\_escfn: \*  
20: em\_escfn: \*  
21: em\_escfn: \*  
22: em\_escfn: \*  
23: em\_escfn: \*  
24: em\_escfn: \*  
25: em\_escfn: \*  
26: em\_escfn: \*  
27: em\_escfn: \*  
28: gb\_esc1: \*

29: gb\_gss2: \*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1598.5	10.3	1076	12	BM480097
2	1578	10.2	896	13	BU173510
3	1523	9.8	907	13	BK409742
4	1477	9.5	900	13	BK460356
5	1464.5	9.4	846	14	CA751074
6	1454.5	9.4	816	13	BU701387
7	1451	9.3	1039	29	CNS0385Y
8	1387	8.9	877	13	BK409741
9	1369	8.8	839	14	CB520719
10	1346	8.7	908	13	BU857213
11	1335.5	8.6	928	12	BT655671
12	1329	8.5	789	14	CB249583
13	1300	8.4	767	14	CB246025
14	1299.5	8.4	893	13	BU849091
15	1284	8.3	812	12	BG915872
16	1281	8.2	694	14	CD350873
17	1267.5	8.2	763	14	CD355642
18	1254	8.1	890	13	BQ920714
19	1240	8.0	745	14	CA328479
20	1180	7.6	598	9	AL704278
21	1180	7.6	719	14	CB245128
22	1173	7.5	597	14	CA891550
23	1173	7.5	723	14	CD350951
24	1170	7.5	738	14	CA510833
25	1168	7.5	833	29	CNS0382Y
26	1166	7.5	722	12	BT456543
27	1141	7.3	875	12	BT693814
28	1121.5	7.2	695	9	AU116945
29	1118	7.2	767	13	BU705756
30	1115	7.2	980	29	CNS04381
31	1114	7.2	662	9	AL042580
32	1101	7.1	861	29	CNS0320M
33	1094	7.0	716	14	CD351314
34	1088.5	7.0	927	13	BQ679970
35	1081.5	7.0	906	13	BQ677400
36	1078	6.9	659	14	BY727867
37	1074	6.9	647	13	BU703750
38	1072.5	6.9	930	13	BQ676596
39	1063	6.8	640	9	AM370778
40	1038	6.7	714	12	BT1078247
41	1035	6.7	4075	11	AK048080
42	1011	6.5	616	9	AW748496
43	1007	6.5	660	10	BE226430
44	1003	6.5	907	14	CA471519
45	997	6.4	1137	10	BF121073

#### ALIGNMENTS

RESULT 1  
LOCUS BM480097 1076 bp mRNA EST 05-FEB-2002  
DEFINITION AGENCOURT 6468579 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5574047  
5' mRNA sequence.  
ACCESSION BM480097  
VERSION BM480097.1 GI:18529139  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 1076)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LHAM12322 row: a column: 24  
 High quality sequence start: 29  
 High quality sequence stop: 666.  
 Location/Qualifiers

FEATURES  
 source 1..1076

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5574047"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_68"  
 /note="Organ: small intestine; Vector: pCMV-SPORTS;  
 Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally;  
 oligo-dt primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH MGC library."  
 BASE COUNT 215 a 357 c 310 g 193 t I others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.02e-92 Length: 1076  
 Score: 1598.50 Matches: 319  
 Percent Similarity: 91.71% Conservative: 13  
 Best local Similarity: 88.12% Mismatches: 18  
 Query Match: 10.28% Indels: 13  
 Gaps: 4

US-09-916-849a-3 (1-2923) x BM480097 (1-1076)

QY 1930 G1ySerleuSerArgValCyAspProGluAspGly-GlnCyAspProGlySerProGluVal 1949  
 Db 1 GGTGGATTTCGGGAATTCGTGACCCCGCTCGATCATGTCATGCAAGCCAGGTGT 60  
 QY 1949 111eG1yArgGlnCyAspArgCyAspAspProPheAlaGluValThrThrAsnGlyCy 1969  
 Db 61 CATGGGCGTCAAGTGTACCCCTGTGACACCTTTTCTGAGTCAACCAATGCTGTG 120  
 QY 1969 eGluValAsnTyAspSerCyProArgAla11eGluAlaGly11eTTPTPProArgTh 1989  
 Db 121 TGAAGTAAATTATACAGCTCCCAACGAGCATTTGAGGTGGATCTGTGGCCCCGTAC 180  
 QY 1989 rArgPheGluValProAlaAlaAlaProCyProGlyGlySerPheGlyThrAlaValAr 2009  
 Db 181 CCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 QY 2009 gH1eCyAspGlu11h1sArgGlyTTPLeuProProAsnLeuPheAsnCyThrSer11eTh 2029  
 Db 241 CCACTGTATAGCAACAGGGGGGTGGCTCCCAAACTCTTCAACTGCAAGTCAATCAC 300  
 QY 2029 rPheSerGluLeuValGlyPheAlaGluValGluGlnArgAsnGluSerGlyLeuAspSe 2049  
 Db 301 CTTCTCAAGACTGAAGGGCTTCTGCTGAGCGGCTACAGCGAATGATGATGAGCTTACG 360  
 QY 2049 rG1yArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTy 2069  
 Db 361 AGGGCGCTCCCAAGAGCTAGCCCTGCTCTCTGCAACCCGACGACACACAGCTGGCTA 420  
 QY 2069 rPheG1ySerAspVal1yValAla1yTg1nLeuAlaThrArgLeuLeuAla1sGluSe 2089  
 Db 421 CTTGGGAGGACACGTCAGAGTGGCTTACAGCTGGCCAGCGGGCTGGGCCCAAGAG 480

QY 2089 rThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnLeu 2109  
 Db 481 CACCCAGCCGGGCTTTGGGCTGTCTCCACACAGACGTCGACTTCACTGAGAAATGTGT 540  
 QY 2109 uArgValGlySerAlaLeuLeuAspThrAlaAsnTyAspGln1sTTPGluLeu11eGln 2129  
 Db 541 GCGGGTGGGACAGGCGCTCTCTGACACAGCCAAAGCGGACCTGAGCTGATTCAGCA 600  
 QY 2129 nThrGlnGlyGlyThAla1ThPheLeuGlnHis1sTyrGluAla1yAla1sSerAlaLeu 2149  
 Db 601 GACAGAGGGTGGACCGCTGCTGCTCAAGCACTATAGAGCTTACGCAAGTCCCTGGC 660  
 QY 2149 ag1nAsnMetArgHis1sThrTyLeuSerProPheThr11eValThProAsn11eVal11 2169  
 Db 661 CAGAACATGCGGACACCTTACCTTAAGCCCTTCAACATGTCAGGCCCAATGTCTCAT 720  
 QY 2169 eSerVal1ValArgLeuAspGlyAsnPheAlaGlyAlaValLeuProArgTyrGluAl 2189  
 Db 721 CTCCTGATGTCGCTGTGACAAAGAACTTGTGGGGCCAAAGTCCCGCTACAGAGGC 780  
 QY 2189 AlaLeuArgGlyGluGlnProProAspLeuGlnThrThrVal11eLeuProGluSerValPh 2209  
 Db 781 CTTGCTGGGAGACACCCCGGACCTTACAGACACAGTCAATTTCTGATGCTTCTT 840  
 QY 2209 eArgGlnThrProProValValArgProAlaGlyProGlyGlu-AlaGlnGluProGlu 2229  
 Db 841 TCAGAGAGACGCGCCCGTGTGTAGGCGCAGGCGCCGAGAAAGGCCAGAGCCACAG 900  
 QY 2229 1u-LeuAlaArgArgGlnArgArg-HisProGluLeuSer---GlnGlyGluAlaVal 2247  
 Db 901 AACCTGGCCCGTGAAGAGACCGGAACCGGAGTTTACCAACCGGGGAGAGCTGTGGC 960  
 QY 2247 aSerVal11e11e-----TyrArgThrLeuAlaGlyLeuLeuProHisAsnTy 2263  
 Db 961 AACGATTCATTCAACCCCAACCCCTGGGGGGGTATTTGCTCTTAACT-----Th 1014  
 QY 2263 rAspProAsp1yAspArgSerLeu--ArgValPro1yArgPro11e11eAsnThrPro 2281  
 Db 1015 TGACCTGAAACAAAC-C-CACTTTGAAAGTCCCAACCGCCGATCATATCAACCC 1071

## RESULT 2

LOCUS BU173510 896 bp mRNA linear EST 04-SEP-2002  
 DEFINITION AGENCOURT 7939348 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6172948  
 5', mRNA sequence.

ACCESSION BU173510.1 GI:22687494  
 VERSION BU173510.1  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 896)  
 TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LHAM1543 row: 1 column: 05  
 High quality sequence stop: 677.  
 Location/Qualifiers

## FEATURES

source 1..896  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

```

/clone="IMAGE:6172948"
/issue_type="retinoblastoma"
/lab_host="DH10B (phage-resilient)"
/clone_lib="NH MGC 67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."
BASE COUNT      169 a      282 c      267 g      177 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      1,66e-91      Length:      896
Score:          1578.00      Matches:      295
Percent Similarity: 98.99%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:    10.15%      Indels:      2
DB:             13      Gaps:      0

US-09-916-849a-3 (1-2923) x BU173510 (1-896)

QY      1351 PhelysCyAspCyProsergylAspPhegluyluylProtyrCyeglinalthrtaryx      1370
DB      1 TTCAAGTGGATGTCCTGAGACTTGGAGACTTGGAGAGCCCTTACTGCGAGGTGACACGCGC      60

QY      1371 SerpPheProAlaHisSerPheIleThrPhearglyLeuargGlnarPheHisPheThr      1390
DB      61 AGCTTCCCGCCCACTCTTCATCACTTTCGGGGCTGGCGAGGTTTCCACTTCAACC      120

QY      1391 LeuAlaLeuSerPheAlaThrlysglyuaargapglyLeuLeuLeuThrArglyarPhe      1410
DB      121 CTGGGCTCTCTCTTGGCCCAAGAGGCGCGAGCGGTTCCTGTGTACATGAGGGGCTTC      180

QY      1411 AengluyluylAspPheValAlaLeuGluValIleGlnGluGlyValGlnLeuThrPhe      1430
DB      181 AATGAGAGCATGACTTGTGGCTTCGAGGTGATTCAGAGAGAGGTCCAGCTCACCTTC      240

QY      1431 SerAlaGlyIuSerThrThrThrValSerProPheValProGlyIValSeraspGly      1450
DB      241 TCTGCAAGGGAGTCAACACCAACGAGGTGCCCATTCGTCGCCGAGAGATCAGTATGGC      300

QY      1451 GlnTPHISThrValGlnLeuLeuTyrrTyraLeuLeuProLeuLeuGlyIValProasp      1470
DB      301 CAGTGCATACAGGTGAGCTGAATACATCAATTAAGCCACTGTGGGTCAACAGGGCTC      360

QY      1471 ProGlnGlyProSerGlyGlnlyValAlaValAlaValThrValAspGlyCyAspThrGly      1490
DB      361 CCAAGGGCCCATCAGAGCAAGAGGTGCTGTGTACCTGTGATGCTGTACACAGGA      420

QY      1491 ValAlaLeuargPheGlySerValleuGlyAsnTyrrSerCyAlaAlaGlnGlyThrGln      1510
DB      421 GTGGCTTGGCTGGATCTGTCTGGGCACTACTCTGTGTGGCCAGGGGCAACCCAG      480

QY      1511 GlyIySerlylylySerleuaspLeuThrGlyProLeuLeuLeuGlyIValProasp      1530
DB      481 GGTGGCAGCAAGAAAGTCTGTGATTCAGAGGGGCCCTCTACTAGCGGGGTGGCTGAC      540

QY      1531 LeuProGluSerPheProValArgMetArgGlnPheValGlyCyMetArgAsnLeuGln      1550
DB      541 CTGGCCGAGAGCTTCCCACTCCGAATCGGCGAGTGTGTGGCTGTCAATGGGAACTTGGAG      600

QY      1551 ValAspSerArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrValProGly      1570
DB      601 GTGGACAGCGGCACTAGACATGGGTGACTTCACTTGCACAAATGGGCAACCGTGGCTGGC      660

QY      1571 CySPeAlaAlaLylyAsnValCyAspSerAsnThrCyHisAsnGlyIlyThrCyVal      1590
DB      661 TCCCTGTGCAAGAAAGCTGTGTGACAGCAACACTTGCACAAATGGGGGCACTTGGCGG      720

QY      1591 AengINTPAspAlaPheSerCyGlyCyProLeuGlyPhe-GlyGlyLylySerCyAla      1610
DB      721 AACGATGGGAGCGCTTCACTGCTGCAAGTCCCTCTGCTGGGGGGCAAGAGCTGGCC      780

```

```

QY      1610 agInGluMetAlaAsnProGlnHisPheLeuGlySerSerleu-ValAlaTPHISgIYL      1630
DB      781 CCAAGAAATGGCCATATCACAGCACTTCTGGGAGGAGCGCTGGGTGGCTGGATGGCC      840

QY      1630 eusSerleuProIleSerGlnProTrrPylLeuSerleuMetPheargThr      1646
DB      841 TCTGCTGCCCATCTCCCAACCCCTGTGATCTCAAGCTCATGTTCGCGAGC      890

RESULT 3
BX409742
LOCUS      BX409742      907 bp      mRNA      linear      EST 13-MAY-2003
DEFINITION      BX409742 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION      CS0DF013YE18 5-PRIME, mRNA sequence.
VERSION      BX409742
KEYWORDS      BX409742.1 GI:30653001
SOURCE      EST.
ORGANISM      Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 907)
AUTHORS      Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 64.f For more
information about this cluster, see http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF012ZG07_Af01112_2&cluster=64.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0BAF012ZG07_Af01112_2.
FEATURES
source
location/Qualifiers
1..907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF013YE18"
/issue_type="FETAL BRAIN"
/dev_stage="fetal"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo (dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and KcoRI sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT      148 a      322 c      238 g      198 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      5.8e-88      Length:      907
Score:          1523.00      Matches:      297
Percent Similarity: 98.67%      Conservative: 0
Best Local Similarity: 98.67%      Mismatches: 4
Query Match:    9.80%      Indels:      2
DB:             13      Gaps:      0

US-09-916-849a-3 (1-2923) x BX409742 (1-907)

QY      2433 AlaAspLeuProPheAlaCyThrValIleAlaIleuLeuHisPheLeuTyrrLeuCyS      2452
DB      7 GCTGACCTCCCTTTTGGCTGCAAGTCAATTCCTGCTGCACTTCTCTACTCTGCGC      66

QY      2453 ThrPheSerTPAlaLeuLeuGluAlaLeuHisLeuTyrrArgAlaLeuThrGluValArg      2472
DB      67 ACTTTTCTGTGGCTGTGCTGAGAGGCTTGACCTGTACCGGGGCACTCACTGAGGTGGCG      126

QY      2473 AspValaLeuThrGlyProMetArgPheTyrrMetLeuGlyIlyTPGlyValProAlaPhe      2492
DB      127 GATGTCAACACCGGCGCCATGCGCTTCACTACATGCTGGGCTGGGGGCTGCTGCTTC      186

```

QY 2493 ILeThrGlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTTP 2512  
 DB 187 ATCAAGAGGCTACCGGCGGCTGAGACCCGAGGGGCTACCGGAACTTCTGCTG 246  
 QY 2513 LeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSer 2532  
 DB 247 CTCCTCCATCTATGACACGCTCATCTGAGATTGCTGGCCGCGTGGCTTTCCTGCTG 306  
 QY 2533 MetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnGlnGly 2552  
 DB 307 ATAGAGTCTTCTGTAACATCTGCGCGCCGCGCTCTGCTGCTGCGCAGCGGAGG 366  
 QY 2553 PheGluValSerGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeu 2572  
 DB 367 TTGAGAGAAAGAGTCTGCTGCTGCGGCTGAGCCCTCTTCCGCTCTGCTGCTG 426  
 QY 2573 SerAlaThrTyrLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheIleTyr 2592  
 DB 427 AGCGCCACGTGGCTGCTGCGACCTGCTCTGTCAACAGCAGACCTCTCTTCCACTAC 486  
 QY 2593 LeuPheAlaThrCysAsnGlyIleGlnGlyProPheIlePheLeuSerTyrValValLeu 2612  
 DB 487 CTCCTTGTCACTGCAATGCAATGCAAGGCGCCCTTCACTCTCTCTATGCTGCTT 546  
 QY 2613 SerIleGluValArgIleValAlaLeuValLeuAlaCysSerArgIleProSerProAspPro 2632  
 DB 547 AGCAAGAGAGTCCGGAAGACATCACTGAGCTGCTGCGACCGCAGCCGACCTGACCT 606  
 QY 2633 AlaLeuThrThrIleSerSerThrIleThrSerSerTyrAsnGlyProSerProTyrAlaAsp 2652  
 DB 607 GCTCTGACCAACCAAGTCACTGAGCTGCTGCTCAACAGTCCGCGCCCTGACGAGAT 666  
 QY 2653 GlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuIleSerThrSerArgSer 2672  
 DB 667 GGGGGGCTGTAACAGCCCTAGAGACTGAGCGGCTCTTGCACAGACCAAGTGGCTG 726  
 QY 2673 GlyIleSerGlnProSerTyrIleProPheLeuLeuArgGluGlnSerAlaLeuAspPro 2692  
 DB 727 GCGAAGAGTCAAGCCGACCTACATCCCTTCTGCTGAGGAGAGTCCGACGAACTT 786  
 QY 2693 GlyIleGlnIleProProGlyLeuGlyAspProGlySerLeuPheLeuGlnGlyIleAspGln 2712  
 DB 787 GGGCAAGGGGCCCCCTGGCTGGGGGATCCAGGCGCTGCTTCTGGGAGGTCAAGACG- 845  
 QY 2713 GlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAspGlnInsery 2732  
 DB 846 CAGCATGATCTGACACGAGCTGACGAGTGAAGTCTTCTTGAAGA- GACCAAGTGGC 904  
 QY 2733 Ser 2733  
 DB 905 TCT 907  
 RESULT 4  
 LOCUS BX460356 990 bp mRNA linear EST 22-MAY-2003  
 DEFINITION BX460356 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 CSODFP013YE18 5-PRIME, mRNA sequence.  
 ACCESSION BX460356  
 VERSION BX460356.1 GI:31035192  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 990)  
 AUTHORS Li, W. B., Gruber, C., Jeejee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 64.f For more  
 information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODFP013BC09QPl&cluster=64.f>. Contact :  
 Feng Liang Email : [liang@life.com](mailto:liang@life.com) URL :  
<http://fulllength.invitrogen.com/invitrogen/1600>  
 Faraday Avenue Genoscope sequence ID : CSODFP013BC09QPl.

FEATURES  
 source  
 1..990  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODFP013YE18"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT\_6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."

BASE COUNT 193 a 318 c 273 g 195 t 11 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,89e-85 Length: 990  
 Score: 1477.00 Matches: 302  
 Percent Similarity: 97.73% Conservative: 0  
 Best Local Similarity: 97.73% Mismatches: 6  
 Query Match: 9.50% Indels: 4  
 DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BX460356 (1-990)  
 QY 2159 ProPheThrIleValThrProAsnIleValIleSerValValArgLeuAspIleGlyAsn 2178  
 DB 64 CCTTCCACCCAGTCAAGCCGCAATTCATCTCCGTAAGTCCCTTGACAAAGGAGAAC 123  
 QY 2179 PheAlaGlyValAlaLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAspLeu 2198  
 DB 124 TTTCCTGGGGCCAGAGTCCCGGCTACAGAGCCCTGCGTGGGAGAGCCCGGACCTT 183  
 QY 2199 GlnThrThrValIleLeuProGluSerValPheArgIleThrProProValValArgPro 2218  
 DB 184 GAGACAAAGTCAATCTGCTGAGTCTGCTTCAAGAGACCCCGCTGAGTCAAGGCC 243  
 QY 2219 AlaGlyProGlyValAlaGlnIleProGluGlnLeuAlaArgGlnArgGlnArgPro 2238  
 DB 244 GCAAGGCCCGAGAGAGCCGAGAGCCAGAGAGCTGGCAAGGCAAGGCAAGCCCG 303  
 QY 2239 GlnLeuSerGlnGlyValAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeu 2258  
 DB 304 GACCTGAGCCAGAGTGAAGGCTGTGGCCAGCTGATCTACCGACCTTGGCCGGGCTA 363  
 QY 2259 LeuProHisAsnTyrAspProAspIleValArgSerLeuArgValProIleIle 2278  
 DB 364 CTGCTCATTAATAGACCTGAGCAAGCGCAGTGA-NGAGTCCCAAGCGCCGATCATC 422  
 QY 2279 AsnThrProValValSerIleSerValHisAspAspGluGlnLeuLeuProArgAlaLeu 2298  
 DB 423 AACACACCCGTGTAGATCAAGCTGCAATGATGATGAGAGACTTCTGCGCGGCGCTG 482  
 QY 2299 AspIleProValThrValGlnPheArgLeuLeuGlnIleGlnIleArgThrIleAspProIle 2318  
 DB 483 GACAAACCGGTACAGGTGAGTTCGGCTGCTGAGAGACAGAGAGGAGCAAGCCATC 542  
 QY 2319 CysValPheTyrAsnHisSerIleLeuValSerGlyThrGlyTyrSerAlaArgGly 2338  
 DB 543 TGTGCTTCTGGAACATTCATCTGCTGAGTGGCAAGGTGGCTGGTGGCCAGAGGC 602  
 QY 2339 CysGluValValPheArgAsnGlnSerHisValSerCysGlnCysAsnHisMetThrSer 2358  
 DB 603 TGTGAAGTGTCTTCCGCAATGAGACAGTCAAGTCTGCAATGACCAATGACAGC 662

QY 2359 PheAlaValIleuMetaspValSerArgArgIuAenGlyGluIle-LeuProIleuYstH 2378  
 DB 663 TTCGGCTGCTCATGACCGTTCTCCGCGGAGAAAGGAGATCCTTCCCACTGAAGAC 722  
 QY 2378 rleuThyTyValAlaLeuGlyValThrlleuAlaIleuLeuLeuThrlPhePhePhe 2398  
 DB 723 ACTGACATACGTCGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 782  
 QY 2398 uThrlleuLeuArgIleLeuArgSerAngIleGlyIleArgArgAsnLeuThrlAlaI 2418  
 DB 783 CACTCTCTCGGATCTCGGCTCCGACCAACGCGCATCCGACCTGACAGCTTC 842  
 QY 2418 AleuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAngIleAlaAspLeuProPheAl 2438  
 DB 843 CCGGGCCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901  
 QY 2438 acyThValIleAlaIleLeuLeuThrlPheLeuThrlLeuCyThThPheSerThrlAlaI 2458  
 DB 902 CTGCAAGTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 QY 2458 uLeuGlyAlaLeuAlaLeuThrlArg 2466  
 DB 961 GCTGGAGGCTTGCACATGTACCG 985  
 RESULT 5  
 CA751074 846 bp mRNA linear EST 27-NOV-2002  
 LOCUS UI-M-FOO-cdm-p-08-0-UI.r1 NIH BMAP\_FOO Mus musculus cDNA clone  
 DEFINITION IMAGE: 6830553 5', mRNA sequence.  
 CA751074  
 VERSION CA751074.1 GI:25577983  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE EST.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 846)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNI at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pyx-5.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..846  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE: 6830553"  
 /rname\_type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /note="Organ: Brain; Vector: pyx-Anc; Site 1: Ecor I;  
 Site 2: Not I; The library was constructed according  
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pyx-Anc vector. The library tag

sequence located between the Not I site and the polyA tail  
 is TCAGAGACC. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

BASE COUNT 172 a 240 c 247 g 184 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,056-84 Length: 846  
 Score: 1464.50 Matches: 258  
 Percent Similarity: 92.61% Conservative: 5  
 Best Local Similarity: 90.85% Mismatches: 19  
 Query Match: 9.42% Indels: 5  
 DB: 14 Gaps: 1

US-09-916-849a-3 (1-2923) x CA751074 (1-846)

QY 1731 GlyProArgLeuHisGlyLeuHisIleu-SerAsnIleThrlValGlyIleProGlyPr 1750  
 DB 3 GGCCCTCGGCTGCACTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 62  
 QY 1750 AlaGlyIleValAlaIleArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspTh 1770  
 DB 63 AGCCAGCGAGAGCTCCGTCGCTCCGCGCTGTCAGAGGTGACAGGAGAGAC 122  
 QY 1770 rProGlyIleValAsnSerLeuAspProSerHisGlyGlnSerIleAsnValGlnGlnG 1790  
 DB 123 ACCTAGAGGTATCAGACACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 182  
 QY 1790 CysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyTyCysSerAs 1810  
 DB 183 CTGTAGCTGCGCAGATCCCTGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 242  
 QY 1810 nAspTrpAspSerTyTySerCysSerCysAspProGlyTyTyTyGlyAspAsnCysThrlAs 1830  
 DB 243 TGACTGGAGACGATATTCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302  
 QY 1830 nValCysAspLeuAsnProCysGlnHisGlnSerValCysThrlArgGlyProSerAlaPr 1850  
 DB 303 TGTGTGATACCTGAAACCATGAGACACAGTCCGTGTGATCCGAAACCATGACAC 362  
 QY 1850 HisGlyTyTyThrlCysGlyCysProProAsnTyTyLeuGlyProTyTyGlyThrlArgG 1870  
 DB 363 CCACGCGTACATCTGCGAGTGTTCACAAATTAATTCCTGCGCATATGTGACACAGAT 422  
 QY 1870 eAspGlnProCysProArgGlyTyTyTyGlyHisProThrlCysGlyProCysAsnCysAs 1890  
 DB 423 TGACCAACCTTGCCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 482  
 QY 1890 rValSerTyGlyPheAspProAspCysAsnTyThrlSerGlyGlyCysHisCysGlyG 1910  
 DB 483 TGTTCAGCAAAAGCTTTGACCCAGATTCGACAAAGCAAGTGGAGGATGTCACAGCAAGA 542  
 QY 1910 uAsnHisTyTyArgProProGlySerProThrlCysLeuLeuCysAspCysTyTyProThrlG 1930  
 DB 543 GAATCACTACCGGCCCCCGGCGAGCCCATCTGTCTGTGTGTGATCTTACCCCATCGG 602  
 QY 1930 ySerLeuSerArgValCysAspProGlyAspGlyGlnCysProCysTyTyProGlyValI 1950  
 DB 603 TCTTTGTCNCAGCTGTGACCCCGAGAGACGCGAGTGTCCGTGCAAGCTCGAGTCA 662  
 QY 1950 eGlyArgGlnCysAspArgCysAspAspProPheAlaGlyValThrlAsnGlyCysG 1970  
 DB 663 TGGGGTCAAGTGTACCGCTGTGACAC-CCTTTGCTAGGTCAACCAATAGCTGTGA 721  
 QY 1970 uValenTyTyAspSerCysProArgAlaIleGlyAlaGlyIleTyTyProArgThrlAr 1990  
 DB 722 AGTGAATTAACACACACTCCACCGGCGCATGAGGCTGGAGTCTGGTGGCCGC-ACCG 780  
 QY 1990 gPheGlyLeuProAlaAlaIleAlaProCysProGlySerPheGlyThrlAlaValArgHis 2010



Db 781 GTTGGGCTA---CTGCTGCTGCCCTGCCCCCANNAGCTCTTGAGACTGTGTGCG-CA 836  
 QY 2010 sCyAaPglu 2013  
 |||||  
 Db 837 CTGTGATGAG 846

RESULT 6  
 BU701387 816 bp mRNA linear EST 09-OCT-2002  
 LOCUS UI-M-EXO-byk-1-14-0-UI.r1 NIH\_BMAP\_EXO Mus musculus cDNA clone  
 DEFINITION IMAGE: 5719381 5', mRNA sequence.

ACCESSION BU701387 GI:23625135  
 VERSION BU701387.1  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 816)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

FEATURES  
 source  
 Location/Qualifiers  
 1..816  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE: 5719381"  
 /issue\_type="whole brain"  
 /dev\_stage="embryo 15.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1ib="NIH BMAP EXO"  
 /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then  
 cloned directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GTGCGTGA. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP): 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institutes of Mental Health  
 (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 164 a 256 c 225 g 168 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,28e-83 Length: 816  
 Score: 1454.50 Matches: 251  
 Percent Similarity: 97.33% Conservative: 4  
 Best Local Similarity: 95.80% Mismatches: 6  
 Query Match: 9.36% Indels: 1  
 DB: 13 Gaps: 1

US-09-916-849a-3 (1-2923) x BU701387 (1-816)

QY 1838 GIuHsgInSerValCysThrArgLysProSerAlaProHsiSglYrThrCysGluCys 1857  
 Db 3 GAGCAGCAGCTCGTGTGTATCCGAAACCCATATACACCCAGCGGTATCGCAGGT 62  
 QY 1858 ProProAntYrLleuGlyProYrCysGluThrArgLleuArgLysProCys---ProArg 1876  
 Db 63 TTACCAAAATTACCTTTGGGCAATATTGTAGACACAGAGATTGACCAACTTTGGCCCCGANN 122  
 QY 1877 GYrTPrPGYrLHieProThrCysGlyProCysAncyAspValSerLysLysPheAsp 1896  
 Db 123 GCGTGTGGGAGACACCCCAATGTGTCTCATACCACTGTATGTACGAAAGGCTTTGAC 182  
 QY 1897 ProAspCysAsnLysThrSerGlyGluCysHsiCysGlyGluAsnHieLysArgProPro 1916  
 Db 183 CCAAGATTGCAACAAAGACAAGTGGGAGTGGCCACTGCAGAGAAATCACTACCGGCCCCC 242  
 QY 1917 GlySerProThrCysLeuLeuCysAspCysYrYrProThrGlySerLeuSerArgValCys 1936  
 Db 243 GGCACGCCCACTTGTCTTGTGTGATACCTGTTACCCACTGTCTTTGTCTTCCGACATCTGT 302  
 QY 1937 AAPProGluAspGlyGluCysProCysLysPProGlyValLleGlyArgGluCysAspArg 1956  
 Db 303 GACCCCGAGAGACGGCCAGTGTCCGTGCAAGCTGAGTCAATTGGGCGTCAGTGTACCC 362  
 QY 1957 CysAspAsnProPheAlaGluValThrThrAsnGlyCysGluValAsnTyraPserCys 1976  
 Db 363 TGTGACCAACCTTTTGTGAGGTACACCAATAGCTGTGAATGAAATTACAGACAGTGC 422  
 QY 1977 ProArgAlaLleGluAlaGlyLleThrPProArgThrArgPheGlyLeuProAlaAla 1996  
 Db 423 CCAAGGCGCATAGAGCTGGAGTCTGGGCCCCCGCCAGCGGTGGCTACCTCTGT 482  
 QY 1997 AlaProCysProLysGlySerPheGlyThrAlaValArgHieCysAspGluHsiArgGly 2016  
 Db 483 GCGCCCTGCCCCAAGAGGCTCTTTGGAGCTGTGTGCGCACTGTGATGACAGCAGGGGT 542  
 QY 2017 TPLeuProProAsnLeuPheAsnCyThrSerLleThrPheSerGluLeuLysGlyPhe 2036  
 Db 543 TGGCTTCCCCCAAACTCTTCACTGACAGTGCACCTTCTCAGAGCTAAAGGGCTTC 602  
 QY 2037 AlaGluValGluAlaArgAsnGluSerGlyLeuAspSerGlyArgSerGluGluLeuAla 2056  
 Db 603 GGTGAGCGGCTGCAGAGAAACAAATCAGCGCTGAGACTGAGCTCCAGAGGGCTAGGC 662  
 QY 2057 LeuLeuLeuArgAsnAlaThrGluHsiThrAlaGlyYrPheGlySerAspValLysVal 2076  
 Db 663 CTGCTCTCGCGAAATGCCACCCAGCACACTCTGCTACTTCGGCAGATGATCAAGGT 722  
 QY 2077 AlaTyrgluLeuAlaThrArgLeuLeuAlaHsiGluSerThrGluArgGlyPheGlyLeu 2096  
 Db 723 GCTTACCAAGCTGGCCACAGCGCTTGTGGCTCATGAGAGTGCACAGCGGGCTTTGGGCTA 782  
 QY 2097 SerAla 2098  
 Db 783 TCCGCA 788

RESULT 7  
 CNS0385Y/c 1039 bp DNA linear GSS 01-SEP-2000  
 LOCUS Tetradodon nigroviridis genome survey sequence T7 end of clone  
 DEFINITION 004013 of library G from Tetradodon nigroviridis, genomic survey  
 sequence.

ACCESSION AL232207  
 VERSION AL232207.1 GI:7891342  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradodon nigroviridis  
 ORGANISM Tetradodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphae; Acanthopterygii; Percomorphae; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Tetradodon.

REFERENCE  
 1 Roest Crolius, H., Jallion, O., Dadiya, C., Bouneau, L., Fisher, C.,

TITLE	JOURNAL	PUBLISHED	REFERENCE	AUTHORS
Bernot, A., Flzames, C., Wincker, P., Brotlier, P., Quetier, F., Saurin, W. and Weissenbach, J.	Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence	Nat. Genet. 25 (2), 235-238 (2000)	10835645	2
Roeet Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C., Flzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis	Genome Res. 10 (7), 939-949 (2000)	20359837	10899143
JOURNAL	20359837	10899143	3 (bases 1 to 1039)	Genoscope.
REFERENCE	Submitted (12-APR-2000)	Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)	Web : www.genoscope.cns.fr	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon.
COMMENT	Location/Qualifiers	1..1039	/organism="Tetradon nigroviridis"	/mol_type="genomic DNA"
FEATURES	source	/db_xref="taxon:99883"	/clone="004J13"	/clone="11b="G"
BASE COUNT	182 a	278 c	329 g	249 t
ORIGIN	1 others			
Alignment Scores:	2.94e-83	Length: 1039	Matches: 277	Conservative: 33
Pred. No.:	1451.00	Matches: 277	Conservative: 33	Mismatches: 32
Score:	90.38%	Matches: 277	Conservative: 33	Mismatches: 32
Percent Similarity:	80.76%	Matches: 277	Conservative: 33	Mismatches: 32
Best local Similarity:	9.33%	Matches: 277	Conservative: 33	Mismatches: 32
Query Match:	29	Gaps: 0		
US-09-916-849A-3 (1-2923) x CNS0385Y (1-1039)				
434 ValHisrYrSerIleMetSerGlyValAsnAlaArgGlyGlnPheTYrLeuAspAlaGlnThr	453			
1034 GTTCACTTCACATCATATGAGCGGCAACACAGGGGCGATTTCATCATGACGCCCAACACC	975			
454 GAlaAlaLeuAspValValSerProLeuAspTYrGluThrThrIlyGluTYrThreularg	473			
974 GCGAAATAGCACTGTGTATGATCCTGTGATCTACGAGGCCAACAAGATTACACCTTGAAG	915			
474 ValArgAlaGlnAspGlyGlyArgProLeuSerAsnValSerGlyLeuValThrVal	493			
914 ATCAGAGCTCAGATGAGAGAGCGGCCCGCTCTCAACATCAGCGCTGTGACGCTG	855			
494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr	513			
854 CAGGTGCTGAGTGTAAACGACACACCCCACTTTGTACAGACGCTTCAGGCCAC	795			
514 ValLeuGlnSerValProLeuGlyTYrTyrLeuValLeuHISValGlnAlaIleAspAlaAsp	533			
794 GTCTCTGAGGAACGTGCCCTGTGTTATCTTCATCATCAATTCAGACGGTGAACGGAC	735			
534 AlaGlyAspAsnAlaArgLeuGlnTYrArgLeuAlaGlyValGlyHISAspPheProPhe	553			
734 TCTGGGGGCAACTCAGGCTGGAAGTACGCTCCTCAACAGCAATTTCCCTTC	675			
554 ThrIleAsnAngLYThrGlyTYrPrlIleSerValAlaAlaGluLeuAspArgGluVal	573			

Db	674	CCGATTCACACACACGCGCTGATGTGTGGCGCGCGCACTGGACCGAGAGAGCCTC	615
Qy	574	AspPheTyrSerPheGlyVal <sup>1</sup> GluAlaArgAspHis <sup>6</sup> GlyThrProAlaLeuThrAlaSer	593
Db	614	GACTTTTACAACTTCGCGGGGTGGAGCGCGGGATCAGGGCTTACCCGTCATCTCCCTCC	555
Qy	594	AlaSerValSerVal <sup>1</sup> ThrValLeuAspVal <sup>1</sup> AsnAspAsnProThrPheThrGlnPro	613
Db	554	GCCAGCATCAGACAGACTGTTTGTGGAGTCAACAGACAAACCCGAGACTCAGCAGAAA	495
Qy	614	GluTyrThrValArgLeuAsnGluAspAla <sup>1</sup> AlaValGlyThrSerVal <sup>1</sup> ThrValSer	633
Db	494	GCTTACTACATACATGGCGCTCAAGAGAGCGCGCGCTGGGGACCAACGCTGTCTACGCTGTG	435
Qy	634	AlaValAspArgAspAla <sup>1</sup> HisSerVal <sup>1</sup> IleThrTyrGlnIleThrSerGlyAsnThrArg	653
Db	434	GCCGTGACCCGAGCAGATCAACAGCGGTGTGACTTATCAGATATTCAGCGGAAACACCGCG	375
Qy	654	AsnArgPheSerIleThrSerGlnSerGly <sup>1</sup> GlyGlyLeu <sup>1</sup> ValSerLeuAlaLeuProLe	673
Db	374	AACGATTTCTCATTTACCCAGCGAGCGCGGTGGCTCATTTTCGCTGGCGCTCTT	315
Qy	673	UAspTyr <sup>1</sup> LeuGln <sup>1</sup> LeuArgGlnTyrVal <sup>1</sup> IleAlaVal <sup>1</sup> ThrAlaSerAspGlyThrArg <sup>1</sup>	693
Db	314	AGACTTACAACTGGAAAGCCCAATATGTCTCATCATCAGCCGACAGTACGSCAGCTCTT	255
Qy	693	NAspThrAlaGlnIleVal <sup>1</sup> ValAsnVal <sup>1</sup> ThrAspAla <sup>1</sup> AsnThrHisArgProVal <sup>1</sup> PheG <sup>1</sup>	713
Db	254	CGACACCGGCTTAAAGTGTGTGTCAACGTCGACCGAATGTCAACACACCGGCCCGGTGTCA	195
Qy	713	NSerSerHisTyr <sup>1</sup> ThrValAsnVal <sup>1</sup> AsnGluAspArgProAlaGly <sup>1</sup> ThrThrValValLe	733
Db	194	GAGCTCATATTACACCGCTCAACATCAACGAAAGCGCGCCATGGGCAACCGGTGGTCT	135
Qy	733	UileSerAlaThrAspGluAspThr <sup>1</sup> GlyGluAsnAla <sup>1</sup> ArgIle <sup>1</sup> ThrTyrPheMetGluAs	753
Db	134	AATTAAGGCCACCGACACAGACACGCGGCGAGAACCGCGCATCCTTAATTATGACAGA	75
Qy	753	PSerIleProGlnPheArgIle <sup>1</sup> AspAlaAspThr <sup>1</sup> GlyAlaVal <sup>1</sup> ThrThrGlnAlaGluLe	773
Db	74	CAGCATCCCTCAGTTCGATATCGAACCTTGACAGCGAGGAGCATGCCACGACGAGATGAGCT	15
Qy	773	UAspTyr <sup>1</sup> 775	
Db	14	GGACTAC 8	
RESULT 8			
BK409741			
LOCUS	BK409741	877 bp	mRNA
DEFINITION	BK409741 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone		EST 13-MAY-2003
VERSION	CS0DF013YE18 5-PRIME, mRNA sequence.		
ACCESSION	BK409741		
VERSION	BK409741.1	GI:30652999	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 877)		
JOURNAL	L.J. M.B., Gruber, C., Jesses, J. and Polyses, D.		
COMMENT	Full-length cDNA libraries and normalization		
	Unpublished		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. This sequence belongs to technology cluster 64.f For more		
	information about this cluster, see http://www.genoscope.cns.fr/		
	cgi-bin/cluster.cgi?seq=CS0BAF012ZG07.AF01112.1&cluster=64.f.		
	Contact : Feng Liang Email : fliang@life tech.com URL :		
	http://fulllength.invitrogen.com/Invitrogen Corporation 1600		
	Faraday Avenue Genoscope sequence ID : CS0BAF012ZG07.AF01112_1.		

FEATURES  
Source

Location/Qualifiers  
1. 877  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DP013YE18"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 140 a 315 c 226 g 194 t 2 others

Alignment Scores:  
Pred. No.: 3.06e-79 Length: 877  
Score: 1387.00 Matches: 275  
Percent Similarity: 94.50% Conservative: 0  
Best Local Similarity: 94.50% Mismatches: 16  
Query Match: 8.92% Indels: 2  
DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BX409741 (1-877)

2433 AlaAspLeuProPheAlaCyThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCyS 2452  
6 GCTACCTCTCTTCTGCTGCTGCA-GTCACTGCCATCTGCTGCACTTCTGCTGCTGCTG 64

2453 ThrPheSerTrpAlaLeuLeuGluAlaLeuHisLeuTyrArgAlaLeuTrpGluValArg 2472  
65 ACCCTTTCCTGGGCTGCTGCTGAGAGCCCTTGACCTGACCGGACCTGACCTGAGTGGC 124

2473 AspValAsnTrpArgPheMetArgPheTyrTyrMetLeuGlyTrpGlyValProAlaPhe 2492  
125 GATGTCAACACGGGCCCCATGCGCTTCTACACATGCTGGGCTGGGCTGGGCTGGCTTC 184

2493 IleThrGlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCyETP 2512  
185 ATCAAGAGGCTAGCGGCTGAGGCTGAGCCCGAGGGCTACGGGAACTGACTTCTGCTGG 244

2513 LeuSerIleTyrAspTrpIleTyrSerPheAlaGlyProValAlaPheAlaValSer 2532  
245 CTCTCCATCTATGACAGCTCATCTGAGGTTTCTGAGCCCGGCTTGGCCGCTCG 304

2533 MetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGly 2552  
305 ATGAGTGTCTTCTGTCATCTGCGGCGCCGCGGCTCTCTGCTGCTGCGCCAGCGGCAAGGC 364

2553 PheGluTyrGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeu 2572  
365 TTTCAGAGAAAGGTCTCTGCTGCGGCTGAGCCCTCTTCCGCTCTCTGCTGCTGCTG 424

2573 SerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspTrpIleLeuPheHisTyr 2592  
425 AGCCCGACGTGGCTGGCTGCTCTCTGTCACAGCGACACCTCTTCTCTCACTAC 484

2593 LeuPheAlaThrCyAsnCyAlaGlnGlyProPheIlePheLeuSerTyrValValLeu 2612  
485 CTCTTCTACCTGCAATGATCAGAGGCCCCCTTCACTCTCTCTGATGAGTCTT 544

2613 SerIleGluValArgAlaLeuLeuLeuAlaCySerSerArgLysProSerProAspPro 2632  
545 AGCAAGAGGATCCGGAAGACCTCAGCTTGCCTGAGCGGCAAGCCAGCCGACCT 604

2633 AlaLeuThrTrpIleSerTrpLeuThrSerSerTyrAsnCyAspProTyrAlaAsp 2652  
605 GCTGTGACCAACCAAGTCCACCTGACCTGCTTCAACAACTGCCACCCCTTACGACAT 664

2653 GlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSer 2672

DB 665 GGGCGGCTGTACCAAGCCCTACAGACTCGCGCGCTCTTGCACAGCACCACTGCTCG 724

2673 GlyIleSerGlnProSerTyrIlePheProPheLeuArgGluIleSerAlaLeuAspPro 2692  
725 GGGAAAGTCAAGCCAGCTACATCCCTTCTGTCTAGAGGAGAGATCCGAACTGAAACT 784

2693 GlyGlnGlyProProGlyLeuGlyAspProGlySerLeuPheLeuGluGlnIleAspGln 2712  
785 GGGCAAGG-CCCCGTGGCTGAGGATTCAGCCAGGCTTGTCTTGGAGGGCAAGACAG 843

2713 GlnHisAspProAspTrpAspSerAspSerAsp 2723

844 CAGCATGATTTCTGTACACACATTCGACAGTAC 876

RESULT 9  
CB520719 839 bp mRNA linear EST 28-MAR-2003  
LOCUS  
DEFINITION  
UT-M-GIO-cej-j-12-0-UT.r1 NIH\_BMAP\_G10 Mus musculus cDNA clone  
IMAGE: 6840397 5', mRNA sequence.

ACCESSION  
CB520719  
VERSION  
CB520719.1 GI:29354074  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 839)  
REFERENCE  
NIH-MGC <http://img.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouse1.html>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
Seq primer: pYX-5.

FEATURES  
Source

Location/Qualifiers  
1. 839  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE: 6840397"  
/issue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lemmon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a NotI site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the NotI site and the polyA tail is AGCGAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 133 a 280 c 233 g 190 t 3 others

ORIGIN

Alignment Scores:  
Pred. No.: 4.15e-78 Length: 839  
Score: 1369.00 Matches: 263



Db 362 CTGCACAGTAATCCCAAGATGGGAGCCCAAGGCTTGGACAGGCCCCCTGGCCAGAGAC 421

Qy 2794 PheGlyThrThrAlaIleuSerSerGlyAseGlyValProGluGluArgGlu 2813

Db 422 TTTGGAGCCACACCAAGAGAGTAGGCAAGGAGCCCTTAGAGAGAGGAGCTGGGAG 481

Qy 2814 AseGlyAsePalaIleuSerSerGlyAseGlyValProGluGluArgGlu 2833

Db 482 AATGAGAGATGCTGCTCGAGAGGGGTCTTAGGCCCCCTTCAAGGCTTCTGCCCAG 541

Qy 2834 ProHileuGlyIleuLeuGlyValGlyValProThrIleuSerGlyValSerSerLeu 2853

Db 542 CTTCAACAAAGGCAATCTTAGAAGAGTGTCTCCCAACCACTAGCAGAGAGACCTTC 601

Qy 2854 LeuAryLeuProleuGluGluGlyValGlyValSerSerGlyValSerSerAseGlyVal 2873

Db 602 CTGGGCTCCCCCTGAGAGCAATGCAAGGCTTTCCTGGGAGCTCTCTGCTAGTGGAG 661

Qy 2874 SerArgGly-GlyProProProArgProProProArgGluSerLeuGluGluLeu 2893

Db 662 AGCCGGGAGCGGCGCCCTCCGCGCCAGCCGCGGAGAGCCTCAAGAGCAGCTGA 721

Qy 2893 nglyAlaMetProIleAlaMetSerIleValIleGly-ThrValaArgGluAseSerG 2913

Db 722 CGGGGTATGCGCATGCGCATGAGCATCAAGGAGGAGGATGAGACTCTGTCAG 781

Qy 2913 1y 2913

Db 782 GG 783

RESULT 11

BI655671 928 bp mRNA linear EST 12-SBP-2001

LOCUS 603283958F1\_NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:5327966 5',

DEFINITION mRNA sequence.

ACCESSION BI655671

VERSION BI655671.1 GI:15569907

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Bukaryoka; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 928)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Princeton University  
Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:  
http://image.liml.gov  
Plate: LIML1832 row: d column: 15  
High quality sequence stop: 769.  
Location/Qualifiers

FEATURES

source

1..928

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="NMRI"

/db\_xref="taxon:10090"

/clone="IMAGE:5327966"

/tissue\_type="tumor, gross tissue"

/dev\_stage="5 months"

/lab\_host="DH10B"

/clone\_id="NCI CGAP Mam4"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Princeton University"

NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."

BASE COUNT 168 a 302 c 239 g 219 t

ORIGIN

Alignment Scores:

Pred. No.: 6,746-76 Length: 928

Score: 1335.50 Matches: 264

Percent Similarity: 85.37% Conservative: 16

Best Local Similarity: 80.49% Mismatches: 29

Query Match: 8,598 Indels: 20

DB: 12 Gaps: 3

US-09-916-849a-3 (1-2923) x BI655671 (1-928)

Qy 2422 AlaGluLeuValPheLeuGlyIleuGluAlaProLeuProPheAlaCysThrVal 2441

Db 2 GCCAGCTGGTCTTCTCTGGGCAATCAAGGCTGACCTTCTTGGCTGTAGAGTC 61

Qy 2442 IleAlaIleLeuLeuHilePheLeuTyrlleuGlyThrPheSerTrpAlaLeuGluAla 2461

Db 62 ATGCTATCTCTGCTGATCTTCTGTACTGTGACCTTCTCTGGGCTGTGAGAGGC 121

Qy 2462 LeuHileuTyrlleuGlyAlaLeuThrGluValArgAspValAseThrGlyProMetArgPhe 2481

Db 122 TTACACTGTACCGGCGCTCAAGAGGTGGCGCAGCTCAATGCAAGTCCCATGCGTTTC 181

Qy 2482 TyrlleuLeuGlyTyrlleuGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAsp 2501

Db 182 TACTACATGCTGGGCGGCTGCTCTGTCTGTATCAACAGGCTGCTGTGGCTTGGAT 241

Qy 2502 ProGluGlyTyrlleuGlyAseProAspPheCysTrpLeuSerIleTyrlleuAspThrLeuIleTrp 2521

Db 242 CGGAGAGGTATAGGAACTTGTGCTGTGCTGTCTGTGTATGATACCTCATTCGG 301

Qy 2522 SerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrlleuAla 2541

Db 302 AGCTTGTGACCAAGTGGCTTGTCTGTGTGATGAGTGTCTGTATCATCTGTGCG 361

Qy 2542 AlaArgAlaSerCysAlaAlaGluGluGlyPheGlyValGlyValProValSerGly 2561

Db 362 GCCCAAGCTCTGTGTGCTGCCAAGCGAGGCTTGGAGAGAGAGGCTGTCTGGGC 421

Qy 2562 LeuGluProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeu 2581

Db 422 CTGGCTCTCTCTTACAGGCTCTCTGTGCTGAGAGGCCAGTGTGCTGCACTGCTC 481

Qy 2582 SerValaSerAspThrLeuLeuPheHileTyrlleuPheAlaThrCysAspCysIleGln 2601

Db 482 TCTGTCAACAGTGAACCTGTCTCTTCCATCACTCTTGTGCTGCTGCAATTGTGTCCAG 541

Qy 2602 GlyProPheIlePheLeuSerTyrlleuValLeuSerGlyValArgGlyAlaLeuGly 2621

Db 542 GGCCCTTATCTTCTCTCTCTAGTGTGTCAAGAGAGGTCCGAGAGACATCCAG 601

Qy 2622 LeuAlaCysSerArgLeuProSerProAlaLeuThrThrIleSerThrLeuThr 2641

Db 602 TTGGCTGAGCGGAGAGCCAGTCCGACCTGTGACCATTAAGCTACCTGAGCC 661

Qy 2642 SerSerTyrlleuGlyProSerProTyrlleuAlaArgGlyValGlyLeuTyrlleuAsp 2661

Db 662 TGTCTTATATCTGCTCCAGCCCTTACAGCAAGGAGGCTGTACAGCTTACAGAGAT 721

Qy 2662 SerAlaGlySerLeuHileSerThrSerArgSerGlyValSerGluProSerTyrlleuPro 2681

Db 722 TCGGCTGCTGCTGTGACAGTCCAGCGCATCCGCAAGAGTCAAGCATATCAT---CC 778

Qy 2682 PheLeuLeuArgGluGluSerAlaLeuAspProGlyGlyProProGlyLeuGlyAsp 2701

Db 779 CTTCCTTGTGAGAGAGTCAAGCTGAA-CTTGGAGAGTTCCTCTGGGAGAGGAA 837

Qy 2702 ProGlySerLeuLeuGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 2721

Db 838 CCAAAGGAGACGCTTGGAAAGGACATCAAG----- 870

Qy 2722 SerAspSerLeuGluAspAspGlnSerGlySerThrAlaSerThrHisSerSerAsp 2741  
Db 871 -----GATATCCGAAACAGCA-----TCTGAACATGATCGTCC 903

Qy 2742 SerGluGluGluGluGluGluGlu 2749  
Db 904 TGAGATGACAAAGCCGCAAGATCA 927

RESULT 12  
CB249583 789 bp mRNA linear EST 14-FEB-2003  
LOCUS CB249583  
DEFINITION U1-M-Exo-b-yl-c-12-0-U1.r1 NIH-BMAP\_EXO Mus musculus cDNA clone  
IMAGE: 5719619 5', mRNA sequence.

ACCESSION CB249583.1 GI:28369156  
VERSION CB249583.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 789)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@dcf-remail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
Seq primer: pYX-5.  
Location/Qualifiers  
1..789  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE: 5719619"  
/cissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction. ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is GTGCGTGGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT 153 a 233 c 246 g 155 t 2 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1 43e-75 Length: 789  
Score: 1329.00 Matches: 239  
Percent Similarity: 94.68% Conservative: 10  
Best Local Similarity: 90.87% Mismatches: 14  
Query Match: 8.55% Indels: 0

DB: 14 Gaps: 0

US-09-916-849a-3 (1-2923) x CB249583 (1-789)

Qy 1562 IleAlaAsnAsnGlyThrValProGlyCySPRoAlaIleValAsnValCyAspSerAsn 1581  
Db 1 ATTGCCAAGATGGACCTGTGCTGCTGCTGCCACCAAGAAAGTCTGTGCAACAGC 60

Qy 1582 ThrCyAsnAsnGlyGlyThrCyValAsnGlnTrpAspAlaPheSerCySPPro 1601  
Db 61 ATTGCCAAGATGGACCTGTGCTGCTGCTGCCACCAAGAAAGTCTGTGCAACAGC 120

Qy 1602 LeuGlyPheGlyGlyIleSerSerCySPProAlaGlnGlnMetAlaAsnProGlnHisPheLeuGly 1621  
Db 121 CTAGGCTTGGGGGCAAGAGCTGCGCCGCAAGAAATGGCCAAACCCGAGGTTCTCGGGC 180

Qy 1622 SerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpIleuSer 1641  
Db 181 AGCAGCTTGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 240

Qy 1642 LeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyArg 1661  
Db 241 CTCATGTTCCGACACGCGCAGGCAATGCGCTGCTGCAAGCGCGTCAACGAGGGCGCC 300

Qy 1662 SerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThrGly 1681  
Db 301 AGCAGCTTGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 360

Qy 1682 LeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrpHis 1701  
Db 361 CTCAGGCTTCTCTGCTGCTGCTTGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 420

Qy 1702 AlaGlnLeuAlaLeuGlyAlaSerGlyIleProGlyHisAlaIleLeuSerPheAspTrp 1721  
Db 421 GCACAGCTGGCAGCTGGGAGCCAGCGGGGCTGGCCATGCAATCTGCTTGTGACTAT 480

Qy 1722 GlyGlnGlnArgAlaGlyIleValSerLeuGlyProArgLeuHisGlyLeuHisLeuSerAsn 1741  
Db 481 GGGCAACAGAAAGGACAGAGGTATCTGGGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGC 540

Qy 1742 IleThrValGlyGlyIleProGlyProAlaGlyValAlaAspGlyPheArgGlyCyS 1761  
Db 541 ATTACAGTGGGGAGGAGTCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGT 600

Qy 1762 LeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSerHis 1781  
Db 601 TTGCAAGGTGTGAGGGTAAGCAGACACCTGAGGGTATCAGCAGTATCCAGCGCGC 660

Qy 1782 GlyGluSerIleAsnValGluGlnGlyCySPProAlaGlnGlnMetAlaAsnPro 1801  
Db 661 GGGAGAGCATCAATGTGAGCAGAGCTGTAGCTGGCCAGATCCCTGTGATCCGATCCA 720

Qy 1802 CySPProAlaAsnSerIleCySPProAlaAspTrpAspSerIleSerCySPPro 1821  
Db 721 TGCCTTACCAACAGTACTGACAGCAATGACGTGGACACTATTCTTGAGCTGTCTCA 780

Qy 1822 GlyIleTrp 1824

Db 781 NCGTACTAT 789

RESULT 13  
CB246025 767 bp mRNA linear EST 12-FEB-2003  
LOCUS CB246025  
DEFINITION U1-M-P00-edu-b-18-0-U1.r1 NIH-BMAP\_P00 Mus musculus cDNA clone  
IMAGE: 6834451 5', mRNA sequence.

ACCESSION CB246025.1 GI:28367669  
VERSION CB246025.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 767)





laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT 179 a 268 c 274 g 172 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.32e-73 Length: 893  
Score: 1299.50 Matches: 254  
Percent Similarity: 73.43% Conservative: 3  
Best Local Similarity: 72.57% Mismatches: 23  
Query Match: 8.36% Indels: 70  
DB: 13 Gaps: 8

US-09-916-849a-3 (1-2923) x BUS49091 (1-893)

QY 1689 leuGlUProglYArGlAlaAenAPGlyAAPTThiHiaAlaGlnLeuAlaGlnYAla 1708  
DB 2 CTGAGCCAGCGCGCGCAATGACGGTGGACACCATGCGCATGCGCATGCGAGCC 61  
QY 1709 SerGlyProglYHiaAlaIleLeuSerPheAPTyGlyGlnGlnAArgAlaGly 1728  
DB 62 AGCGGGGGCGCGCGCATGCTCTCTCTGATTATGGCGAGAGAGAGAGAGAGG 121  
QY 1729 AenLeuGlyProArGlLeuHiaGlyLeuHiaLeuSerAenIleThyAlaGlyIlePro 1748  
DB 122 AACCTGGCGCGCGCGCGCATGCTGACCTAGACAACTAACAGTGGCGGCAATACCT 181  
QY 1749 GlyProAlaGlyAlaAlaArgGlyPheArGlyCyleuGlnGlnYAlaArgValSer 1768  
DB 182 GGGCCAGCGCGCGCGCGCGCTTCTGGCGCTTCTGGCGCTTCTGGCGCTGAGC 241  
QY 1769 AspThProglYAlaAenSerLeuAenProSerHiaGlyGlnSerIleAenValGlu 1788  
DB 242 GATACCGCGCGCGCGCGTAAACCTGATCCAGCCATGGAGAGAGAGAGAGAG 301  
QY 1789 GlnGlyCySerLeuProAPProCyAAsPSeAenProCyAProAlaAenSerTyCyS 1808  
DB 302 CAAGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
QY 1809 SerAenAPThProAPSerTySerCySerCyAAsPProglYTyTyTyTyTyTyTyTy 1828  
DB 362 AGCAAGACCTGAG 421  
QY 1829 ThrAenValCyAAsPLeuAenProCyGlnHiaGlnSerValCyThArGlySProSer 1848  
DB 422 ACTAATGTGTGTGACCTGAAACCTGTGAGACACAGTGTGTGTGTGTGTGTGTGT 481  
QY 1849 AlaProHiaGlyTy 1868  
DB 482 GCGCCCGCATGCTATACCTGCGAGTGTCTCCCAATTAACCTTGGCCATATCTGTAGAC 541  
QY 1869 ArgIleAAsGlnProCyS-ProArGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 1888  
DB 542 AGGATTTGACCAAGCTTGTCTCCCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 601  
QY 1888 snCyAAsPValSerTyGly-PheAP-ProAPProCyAAsnTyThSerGlyGlnTyCyS 1907  
DB 602 ACTGTATGTCAAGCAAGCTTGTGAACCACTGACACACACACACACACACACACAC 661  
QY 1907 sCyAlTyGlnAenHia-Ty 1926  
DB 662 CTGCAAG 720  
QY 1926 yetyTyProthGlySerLeuSerArGlyValCyAAsPProglYArGlyGlnTyCySProCy 1946  
DB 721 -----ACTGC-----TAACCCCGCGAGGCTTCTTGTGTCA 754  
QY 1946 yS-----ProGlyValIleGlyArGlnTyCyAAsPArGlyAAsPArProPheA 1962  
DB 755 AAATCTGGAG 783

QY 1962 laeGluValThrThraGlnGlyCyGluValAenTyAAsPSeCySProArGAlaIleGluA 1982  
DB 783 ----- 783  
QY 1982 laeGlyIleTy 2002  
DB 784 -----CCCGAGGCTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 814  
QY 2002 TySerPheGlyThraAlaValArGlnHiaGlyAAsPProglYTyTyTyTyTyTyTyTy 2019  
DB 815 GTTCTCTCCGGGGGCT-----CACAGAGGGAGAGAGAGAGAGAGAGAGAGAGAG 856  
QY 2020 -----ProAenLeuPhe 2023  
DB 857 AAACAAACCTCTCTTTT 876

RESULT 15  
BG915872 812 bp mRNA linear EST 05-JUN-2001  
LOCUS 602815323P1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4937669 5',  
DEFINITION mRNA sequence.  
ACCESSION BG915872 GI:14296348  
VERSION BG915872.1 GI:14296348  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 812)  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL  
COMMENT Contract: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch  
Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LMNL0871 row: n column: 06  
High quality sequence stop: 778.  
Location/Qualifiers  
1. 812  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4937669"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_id="NCI CGAP Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furch,  
NIH Reference for transgenic model; Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

#### FEATURES

source

BASE COUNT 165 a 255 c 232 g 160 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.16e-72 Length: 812  
Score: 1284.00 Matches: 254  
Percent Similarity: 94.85% Conservative: 4  
Best Local Similarity: 93.38% Mismatches: 12  
Query Match: 8.26% Indels: 8  
DB: 12 Gaps: 0

US-09-916-849a-3 (1-2923) x BG915872 (1-812)

QY 1894 GlyPheASPProASPcysasnlysrhSerglyglucysHisCyslysgluasnHis-Ty 1913  
DB 2 GGCTTGACCCAGATTGCAACAGACAGTGGGAGTGCACCTGCAGAGGAGATCACTAG 61  
QY 1913 rArgProPArglySerProThrCysleuLeuCysaspCysTyProThrGlySerleuSe 1933  
DB 62 CCGGTCCCGGCGAGCCCACTGTCTCTGTGTGACTGTACCCCACTGGTCTTTGTC 121  
QY 1933 rArgValCysaspProglunaspGlyGlnCysProCyslyPArglyValIleGlyValArgI 1953  
DB 122 CCGAGCTGTGAGCCCGAGAGACCGCCAGTGTCCGTGCAGACCTGCAGAGTCAITGGGCTCA 181  
QY 1953 nCysaspArgCysaspAspProPhealaglValIthrThrAsnGlyCysgluValAsnTy 1973  
DB 182 GTGTGACCGCTGACAACTTTGTGTGAGTCAACCAACATGCGCTGGAAGTGAATTA 241  
QY 1973 rAspSerCysProArgAlallegluagllyleTPTTPProArgThrArgPheGlyle 1993  
DB 242 CGACAGCTGCCCAAGGAGCTAGAGGCTGGATCTGGTGGCCCGCAAGGCTTGGGCT 301  
QY 1993 uProAlaAlaAlaProCysProlysglySerPheGlyThrAlaValArgHisCysaspG 2013  
DB 302 ACCTGCTGCTGCCCTGCCCAAGGCTCTTTGGGACTGTGTGCGCCACTGTGATGA 361  
QY 2013 uHisArgGlyTyrPleuProPheasnleuPheasnCysThrSerIleThrPheSergIle 2033  
DB 362 GCACAGGGGTTGGCTCCCGCAACCTTCACACTGCACGTCAGTCACTTCAGAGCT 421  
QY 2033 uLysglyPheAlaglunargleuGlnargasnGlyleuaspSerglyVarSergI 2053  
DB 422 AAAGGCTTTCGTGAGGAGCTGCAAGAGAAAGATCAGCTGAGTCAAGAGCTTCCA 481  
QY 2053 ngInleuAlaleuLeuLeuArgAsnAlaThrGlnHisThrAlaglYTrPheGlySerAs 2073  
DB 482 GAGGCTAGCCCTGCTCTGCGCAATGCAACCAAGCAACCTTGCGCTAGTGGCACTGA 541  
QY 2073 pValIyvalAlaIlyrGlnleuAlaThrArgleuLeuAlaHisGlySerThrGlnArgI 2093  
DB 542 TGTCAAGGTGGCTTACCAAGCTGCGCAACGCTCTTGGCTCATGAGAGTGGCCAGCGGG 601  
QY 2093 yPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnleuLeuArgValGlySe 2113  
DB 602 CTTTGGGCTATCCGCAACAGATGTACCTTCACGAGAAATCTGCTGAGGTTGGGCA 661  
QY 2113 rAlaleuLeuaspThrAlaAsnlysaArgHisrTPGlnleuIleGlnIthrGlnGlyG 2133  
DB 662 CGC-CTCCTGTGATGC-AGCAACAAGAGGCACTGGAACTGATCCAGCAGACAGAGGTTG 719  
QY 2133 yThrAlaIlePleuLeuGlnHis-TyrGlnuAlaTyrAlaSerAlaLeuAlaGlnAsnMetA 2153  
DB 720 CAGGCC-TGGCTGCTCCAGACATAAGAGGCTTACGCAAGCGC-CTTGCCCAAGATATGC 777  
QY 2153 rHisThrTyrLeuSerProPheThrIleVal 2163  
DB 778 GGCA-AACTACTTAC-CCCTTACCATCGTC 807

Search completed: February 14, 2004, 05:36:28  
Job time : 11547 secs